REPORT DOCUMENTATION PAGE

Form Approved OMB NO. 0704-0188

1. REPORT DATE (DD-MM-YYYY)	2. REPORT TYPE	3. DATES COVERED (From - To)
	New Reprint	-

4. TITLE AND SUBTITLE	
A Lossy Source Coding Interpretation of Wyner's Common Information	5a. CONTRACT NUMBER W911NF-12-1-0383
	5b. GRANT NUMBER
	5c. PROGRAM ELEMENT NUMBER 611102
6. AUTHORS Wei Liu, Biao Chen, Ge Xu	5d. PROJECT NUMBER
	5e. TASK NUMBER
	5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAMES AND ADDRESSES	8. PERFORMING ORGANIZATION REPORT NUMBER
Syracuse University Office of Research 113 Bowne Hall Syracuse, NY 13244 -1200	

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS (ES)	10. SPONSOR/MONITOR'S ACRONYM(S)
U.S. Army Research Office P.O. Box 12211 Research Triangle Park, NC 27709-2211	ARO
	11. SPONSOR/MONITOR'S REPORT NUMBER(S) 61953-CS.13

12. DISTRIBUTION AVAILIBILITY STATEMENT

Approved for public release; distribution is unlimited.

13. SUPPLEMENTARY NOTES

The views, opinions and/or findings contained in this report are those of the author(s) and should not contrued as an official Department of the Army position, policy or decision, unless so designated by other documentation.

14. ABSTRACT

Wyner's common information was originally defined for a pair of dependent discrete random variables. Its significance is largely reflected in, and also confined to, several existing interpretations in various source coding problems. This paper attempts to expand its practical significance by providing a new operational interpretation. In the context of the Gray–Wyner network, it is established that Wyner's common information has a new lossy source coding interpretation. Specifically, it is established that, under suitable conditions, Wyner's common information

15. SUBJECT TERMS

Common information, Gray-Wyner network, rate distortion function

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	15. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			Biao Chen
UU	UU	UU	UU		19b. TELEPHONE NUMBER 315-443-3332

Standard Form 298 (Rev 8/98)
Prescribed by ANSI Std. Z39.18

Report Title

A Lossy Source Coding Interpretation of Wyner's Common Information

ABSTRACT

Wyner's common information was originally defined for a pair of dependent discrete random variables. Its significance is largely reflected in, and also confined to, several existing interpretations in various source coding problems. This paper attempts to expand its practical significance by providing a new operational interpretation. In the context of the Gray–Wyner network, it is established that Wyner's common information has a new lossy source coding interpretation. Specifically, it is established that, under suitable conditions, Wyner's common information equals to the smallest common message rate when the total rate is arbitrarily close to the rate distortion function with joint decoding for the Gray–Wyner network. A surprising
observation is that such equality holds independent of the values of distortion constraints as long as the distortions are within some distortion region. The new lossy source coding interpretation provides the first meaningful justification for defining Wyner's common information for continuous random variables and the result can also be extended to that of multiple variables. Examples are given for characterizing the rate distortion region for the Gray–Wyner lossy source coding problem and for identifying conditions under which Wyner's common information equals that of the smallest common rate. As a by-product, the explicit
expression for the common information between a pair of Gaussian random variables is obtained.

REPORT DOCUMENTATION PAGE (SF298)
(Continuation Sheet)

Continuation for Block 13

ARO Report Number 61953.13-CS
A Lossy Source Coding Interpretation of Wyner'...

Block 13: Supplementary Note

A Lossy Source Coding Interpretation of Wyner's Common Information

Ge Xu, Wei Liu, and Biao Chen, *Fellow, IEEE*

***Abstract*— Wyner's common information was originally defined for a pair of dependent discrete random variables. Its significance is largely reflected in, and also confined to, several existing interpretations in various source coding problems. This paper attempts to expand its practical significance by providing a new operational interpretation. In the context of the Gray–Wyner network, it is established that Wyner's common information has a new lossy source coding interpretation. Specifically, it is established that, under suitable conditions, Wyner's common information equals to the smallest common message rate when the total rate is arbitrarily close to the rate distortion function with joint decoding for the Gray–Wyner network. A surprising observation is that such equality holds independent of the values of distortion constraints as long as the distortions are within some distortion region. The new lossy source coding interpretation provides the first meaningful justification for defining Wyner's common information for continuous random variables and the result can also be extended to that of multiple variables. Examples are given for characterizing the rate distortion region for the Gray–Wyner lossy source coding problem and for identifying conditions under which Wyner's common information equals that of the smallest common rate. As a by-product, the explicit expression for the common information between a pair of Gaussian random variables is obtained.**

***Index Terms*— Common information, Gray-Wyner network, rate distortion function.**

I. Introduction

CONSIDER A pair of dependent random variables X and Y with joint distribution $p(x, y)$, which denotes either the probability density function if X and Y are continuous or the probability mass function if X and Y are discrete. Quantifying the information that is common between X and Y has been a classical problem both in information theory and in mathematical statistics [1]–[4]. The most widely used notion is Shannon's mutual information, defined as

$$I(X; Y) = E\left[\log \frac{p(x, y)}{p(x)p(y)}\right]$$

where $p(x)$ and $p(y)$ are the marginal distributions of X and Y corresponding to the joint distribution $p(x, y)$ and $E[\cdot]$ denotes expectation with respect to $p(x, y)$. Shannon's mutual information measures the amount of uncertainty reduction in one variable by observing the other. Its significance lies in its applications to a broad range of problems in which concrete operational meanings of $I(X; Y)$ can be established. These include both source and channel coding problems in information and communication theory [5] and hypothesis testing problems in statistical inference [6].

Other notions of information have also been defined between a pair of dependent variables. Most notable among them are Gács and Körner's common randomness $K(X, Y)$ [2] and Wyner's common information $C(X, Y)$ [4]. Gács and Körner's common randomness is defined as the maximum number of common bits per symbol that can be independently extracted from X and Y. Quite naturally, $K(X, Y)$ has found extensive applications in secure communications, e.g., for key generation [7]–[9]. More recently, a new interpretation of $K(X, Y)$ using the Gray-Wyner source coding network was given in [10]. It was noted in [2] and [11] that the definition of $K(X, Y)$ is rather restrictive in that $K(X, Y)$ equals 0 in most cases except for the special case when $X = (X', V)$ and $Y = (Y', V)$ and X', Y', V are independent variables or those (X, Y) pair that can be converted to such a dependence structure through relabeling the realizations, i.e., permutation of the joint distribution matrix. Notice also that $K(X, Y)$ is defined only for discrete random variables.

Wyner's common information was originally defined for a pair of discrete random variables with finite alphabet as

$$C(X, Y) = \inf_{X-W-Y} I(X, Y; W). \tag{1}$$

Here, the infimum is taken over all auxiliary random variables W such that X, W, and Y form a Markov chain. The operational meanings of $C(X, Y)$ available in existing literature include the minimum common rate for the Gray-Wyner lossless source coding problem under a sum rate constraint, the minimum rate of a common input of two independent random channels for distribution approximation [4], and strong coordination capacity of a two-node network without common randomness and with actions assigned at one node [12].

While Wyner only considered the definition of common information for discrete random variables, the expression specified by (1) directly applies to a pair of continuous

Manuscript received January 8, 2013; revised February 9, 2015; accepted September 30, 2015. Date of publication December 8, 2015; date of current version January 18, 2016. This work was supported in part by the U.S. Army Research Office under Award W911NF-12-1-0383, in part by the Air Force Office of Scientific Research under Award FA9550-10-1-0458, and in part by the National Science Foundation under Award 1218289. This paper was presented at the 2010 Annual Allerton Conference on Communication, Control, and Computing and the 2011 Annual Conference on Information Sciences and Systems.

G. Xu was with the Department of Electrical Engineering and Computer Science, Syracuse University, Syracuse, NY 13244 USA. She is now with Nuance Communications, Burlington, MA 01803 USA (e-mail: ge.xu@nuance.com).

W. Liu was with the Department of Electrical Engineering and Computer Science, Syracuse University, Syracuse, NY 13244 USA. He is now with Bloomberg L.P., New York, NY 10022 USA (e-mail: wliusyr@gmail.com).

B. Chen is with the Department of Electrical Engineering and Computer Science, Syracuse University, Syracuse, NY 13244 USA (e-mail: bichen@syr.edu).

Communicated by Y. Oohama, Associate Editor for Source Coding.

Digital Object Identifier 10.1109/TIT.2015.2506560

random variables. However, absent of any meaningful interpretation, the computed value for a pair of continuous random variables is largely pointless. Notice that Wyner's original interpretations for $C(X, Y)$ are only applicable to discrete random variables.

This paper presents a new lossy source coding interpretation of Wyner's common information using the Gray-Wyner network. Specifically, we show that, for the Gray-Wyner network, Wyner's common information is precisely the smallest common message rate for a certain range of distortion constraints when the total rate is arbitrarily close to the rate distortion function with joint decoding. As the common information is only a function of the joint distribution, this smallest common rate remains constant even if the distortion constraints vary, as long as they are within a specific distortion region. With this new interpretation, concrete practical interpretation is thus associated with Wyner's common information defined for a pair of continuous random variables. There has also been recent effort in characterizing the common message rate for lossy source coding using the Gray-Wyner network [16]. We establish the equivalence between the characterization in [16] with an alternative characterization presented in the present paper.

Computing Wyner's common information is known to be a challenging problem; $C(X, Y)$ was only resolved for several special cases as described in [4] and [13]. Along with our generalizations of Wyner's common information, we provide two new examples where we can explicitly evaluate the common information of multiple dependent variables. In particular, we derive, through an estimation theoretic approach, $C(X, Y)$ for a bivariate Gaussian source and its extension to the multivariate case with a certain correlation structure.

The rest of the paper is organized as follows. Section II reviews Wyner's two approaches for the common information of two discrete random variables, the marginal, and conditional rate distortion functions and their relationships and the concept of successively refinable sources. In Section III, we provide a new interpretation of Wyner's common information using Gray-Wyner's lossy source coding network. Specifically, we prove that for the Gray-Wyner network, Wyner's common information is precisely the smallest common message rate for a certain range of distortion constraints when the total rate is arbitrarily close to the rate distortion function with joint decoding. In Section IV, two examples, the doubly symmetric binary source and the bivariate Gaussian source, are used to illustrate the lossy source coding interpretation of Wyner's common information. The common information for bivariate Gaussian source and its extension to the multi-variate case is also derived in IV. Section V concludes this paper.

Notation: Throughout this paper, we use calligraphic letter $\mathcal{X}$ to denote the alphabet and $p(x)$ to denote either point mass function or probability density function of a random variable X.

II. Existing Results

A. Wyner's Result

Given two discrete random variables X_1 and X_2 with distribution $p(x_1, x_2)$, Wyner defined the common information of them as in equation (1) and provided two operational meanings for this definition.

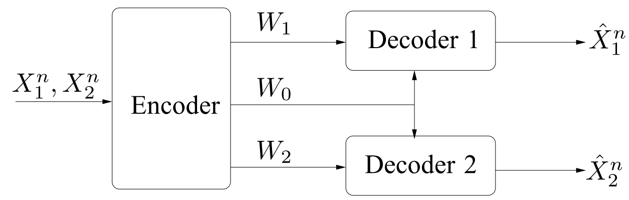

Fig. 1. Source coding over a simple network.

The first approach is based on the model shown in Fig. 1 which is a source coding network first studied by Gray and Wyner in [17]. In this model, the encoder observes a pair of sequences (X_1^n, X_2^n). There are 2 receivers, with the ith receiver only interested in recovering the sequence X_i^n, $i = 1, 2$. The encoder encodes the source into 3 messages W_0, W_1, W_2, one is a public message available at all receivers while the other 2 messages are private messages only available at the corresponding receivers. For $m = 1, 2, \cdots$, let $I_m = \{0, 1, 2, \cdots, m-1\}$. An (n, M_0, M_1, M_2) code is defined as below.

Definition 1: An (n, M_0, M_1, M_2) code consists of

- An encoder mapping

$$f : \mathcal{X}_1^n \times \mathcal{X}_2^n \to I_{M_0} \times I_{M_1} \times I_{M_2},$$

- 2 decoder mappings

$$g_i : I_{M_i} \times I_{M_0} \to \hat{\mathcal{X}}_i^n, \quad i = 1, 2.$$

For an (n, M_0, M_1, M_2) code, let $f(X_1^n, X_2^n) = (W_0, W_1, W_2)$ and $\hat{X}_i^n = g_i(W_i, W_0)$, $i = 1, 2$. The probability of error can be obtained by

$$P_e^{(n)} = \frac{1}{2n} \sum_{i=1}^{2} E[d_H(X_i^n, \hat{X}_i^n)], \tag{2}$$

where $d_H(u^n, \hat{u}^n)$ is the Hamming distance between u^n and $\hat{u}^n$.

A number R_0 is said to be *achievable* if for any $\epsilon > 0$, there exists, for n sufficiently large, an (n, M_0, M_1, M_2) code such that

$$M_0 \le 2^{nR_0}, \tag{3}$$

$$\sum_{i=0}^{2} \frac{1}{n} \log M_i \le H(X_1, X_2) + \epsilon, \tag{4}$$

$$P_e^{(n)} \le \epsilon. \tag{5}$$

Wyner defined C_1 as the the infimum of all achievable R_0 and proved in [4] that

$$C_1 = C(X_1, X_2). \tag{6}$$

The second approach is shown in Fig. 2. In this approach, the joint distribution $p(x_1^n, x_2^n) = \prod_{i=1}^{n} p(x_{1i}, x_{2i})$ is approximated by the output distribution of a pair of random number generators. A common input W, uniformly distributed on $\mathcal{W} = \{1, \cdots, 2^{nR_0}\}$ is sent to two separate processors which are independent of each other. These processors (random number generators) generate independent and

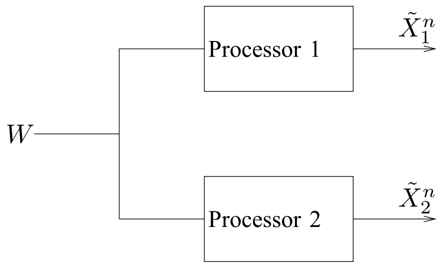

Fig. 2. Random variables generator.

identically distributed (i.i.d) sequences according to two distributions $q_1(x_1^n|w)$ and $q_2(x_2^n|w)$ respectively. The output sequences of the two processors are denoted by $\tilde{X}_1^n$ and $\tilde{X}_2^n$ respectively and the joint distribution of the output sequences is given by

$$q(x_1^n, x_2^n) = \sum_{w \in \mathcal{W}} \frac{1}{|\mathcal{W}|} q_1(x_1^n|w) q_2(x_2^n|w).$$

Let

$$D_n(q, p) = \frac{1}{n} \sum_{x_1^n \in \mathcal{X}_1^n, x_2^n \in \mathcal{X}_2^n} q(x_1^n, x_2^n) \log \frac{q(x_1^n, x_2^n)}{p(x_1^n, x_2^n)}.$$

Let C_2 be the infimum of rate R_0 for the common input such that for any $\epsilon > 0$, there exists a pair of distribtions $q_1(x_1^n|w)$, $q_2(x_2^n|w)$ and n such that $D_n(q, p) \leq \epsilon$.

It is proved in [4] that

$$C_2 = C(X_1, X_2). \tag{7}$$

It is worth noting that Wyner's common information can be generalized to that of multiple dependent random variables by replacing the Markov condition with a conditional independence structure; the latter is equivalent to the Markov condition for three random variables yet is applicable to multiple random variables. As established in [31], such a generalization is meaningful in that both interpretations C_1 and C_2 carry over to the case involving multiple random variables.

B. Joint, Marginal and Conditional Rate Distortion Functions

In this section, we review the joint, marginal and conditional rate distortion functions and their relations which will be used in the following sections. Two-dimensional sources will be considered and the results can be generalized immediately to N-dimensional vector sources.

Given a two-dimensional source (X_1, X_2) with probability distribution $p(x_1, x_2)$ and two distortion measures $d_1(x_1, \hat{x}_1)$ and $d_2(x_2, \hat{x}_2)$ defined on $\mathcal{X}_1 \times \hat{\mathcal{X}}_1$ and $\mathcal{X}_2 \times \hat{\mathcal{X}}_2$, the joint rate distortion function is given by

$$R_{X_1X_2}(D_1, D_2) = \inf_{p(\hat{x}_1\hat{x}_2|x_1x_2)} I(X_1X_2; \hat{X}_1\hat{X}_2), \tag{8}$$

where the minimum is taken over all $p(\hat{x}_1\hat{x}_2|x_1x_2)$ such that $Ed_1(X_1, \hat{X}_1) \leq D_1,\ Ed_2(X_2, \hat{X}_2) \leq D_2$.

The conditional rate distortion function is defined as [20]

$$R_{X_1|X_2}(D_1) = \inf_{p(\hat{x}_1|x_1x_2)} I(X_1; \hat{X}_1|X_2). \tag{9}$$

where the minimum is taken over all $p(\hat{x}_1|x_1, x_2)$ satisfying $Ed_1(X_1, \hat{X}_1) \leq D_1$. The joint, marginal and conditional rate distortion functions satisfy the following inequalities.

Lemma 1 [18], [19]: Given a two-dimensional source (X_1, X_2) with joint distribution $p(x_1, x_2)$ and two distortion measures $d_1(x_1, \hat{x}_1)$, $d_2(x_2, \hat{x}_2)$ defined respectively on $\mathcal{X}_1 \times \hat{\mathcal{X}}_1$ and $\mathcal{X}_2 \times \hat{\mathcal{X}}_2$, the rate distortion functions satisfy the following inequalities

$$R_{X_1X_2}(D_1, D_2) \geq R_{X_1|X_2}(D_1) + R_{X_2}(D_2), \tag{10a}$$

$$R_{X_1|X_2}(D_1) \geq R_{X_1}(D_1) - I(X_1; X_2), \tag{10b}$$

$$R_{X_1X_2}(D_1, D_2) \geq R_{X_1}(D_1) + R_{X_1}(D_2) - I(X_1; X_2), \tag{10c}$$

$$R_{X_1}(D_1) \geq R_{X_1|X_2}(D_1), \tag{11a}$$

$$R_{X_1}(D_1) + R_{X_2}(D_2) \geq R_{X_1X_2}(D_1, D_2). \tag{11b}$$

A sufficient condition for equality in (10) is that the optimum backward test channels for the functions on the left side of each equation factor appropriately, i.e., for (10a) $p_b(x_1x_2|\hat{x}_1\hat{x}_2) = p(x_1|\hat{x}_1x_2)p(x_2|\hat{x}_2)$, for (10b) $p_b(x_1|\hat{x}_1x_2) = p(x_1|\hat{x}_1)$ and for (10c) that $p_b(x_1x_2|\hat{x}_1\hat{x}_2) = p(x_1|\hat{x}_1)p(x_2|\hat{x}_2)$. Equalities hold in (11) if and only if X_1 and X_2 are independent.

It is shown in [18] that under quite general conditions, equalities in (10) hold for small values of distortion. To state this result, we introduce the Extended Shannon Lower Bounds (ESLB) of rate distortion functions [18], [20]. Let us denote $R_U^{(L)}(D)$ the ESLB of $R_U(D)$. $R_U^{(L)}(D)$ is a lower bound of $R_U(D)$ and is derived by removing the constraint that the minimizing test channel $p_t(\hat{u}|u)$ is nonnegative in the minimization defining $R_U(D)$. The detailed definition and calculation of ESLB can be found in [20].

If the reproducing alphabet is identical to the source alphabet, the marginal, joint and conditional rate distortion functions equal to their corresponding ESLB for small distortion regions, while these ESLB, satisfy the equalities in (10). The detail of this result is given in Lemma 2 and the following notations are used. $\mathcal{D}$ denotes a surface in an m-dimensional space and the inequality $\Delta \leq \mathcal{D}$ means that there exists a vector $\beta \in \mathcal{D}$ such that $\Delta \leq \beta$. If there is no such a vector, $\Delta > \mathcal{D}$. Likewise, $\mathcal{D}_1 \leq \mathcal{D}_2$ means that $\beta \leq \mathcal{D}_2$ for any $\beta \in \mathcal{D}_1$ [18].

Lemma 2 [18]: Given a two-dimensional source (X_1, X_2) with joint distribution $p(x_1, x_2)$ and alphabet $\mathcal{X}_1 \times \mathcal{X}_2$, reproduction alphabets $\hat{\mathcal{X}}_1 = \mathcal{X}_1$, $\hat{\mathcal{X}}_2 = \mathcal{X}_2$ and two per-letter distortion measures $d_1(x_1, \hat{x}_1)$ and $d_2(x_2, \hat{x}_2)$ that satisfy

$$d_i(x_i, \hat{x}_i) > d_i(x_i, x_i) = 0, \quad x_i \neq \hat{x}_i,\ i = 1, 2, \tag{12}$$

there exist strictly positive surfaces $\mathcal{D}(X_1X_2)$, $\mathcal{D}(X_1|X_2)$, $\mathcal{D}(X_1)$ and $\mathcal{D}(X_2)$ such that

$$\begin{aligned} R_{X_1X_2}(D_1, D_2) &= R^{(L)}_{X_1X_2}(D_1, D_2), \quad (D_1, D_2) \leq \mathcal{D}(X_1X_2), \\ R_{X_1|X_2}(D_1) &= R^{(L)}_{X_1|X_2}(D_1), \quad D_1 \leq \mathcal{D}(X_1|X_2), \\ R_{X_1}(D_1) &= R^{(L)}_{X_1}(D_1), \quad D_1 \leq \mathcal{D}(X_1), \\ R_{X_2}(D_2) &= R^{(L)}_{X_2}(D_2), \quad D_2 \leq \mathcal{D}(X_2), \end{aligned} \tag{13}$$

and

$$\mathcal{D}(X_1|X_2) \leq \mathcal{D}(X_1), \tag{14}$$

$$\mathcal{D}(X_1X_2) \leq (\mathcal{D}(X_1|X_2), \mathcal{D}(X_2)) \leq (\mathcal{D}(X_1), \mathcal{D}(X_2)). \tag{15}$$

Finally,

$$R^{(L)}_{X_1X_2}(D_1, D_2) = R^{(L)}_{X_1|X_2}(D_1) + R^{(L)}_{X_2}(D_2), \tag{16}$$

$$R^{(L)}_{X_1|X_2}(D_1) = R^{(L)}_{X_1}(D_1) - I(X_1; X_2), \tag{17}$$

$$R^{(L)}_{X_1X_2}(D_1, D_2) = R^{(L)}_{X_1}(D_1) + R^{(L)}_{X_2}(D_2) - I(X_1; X_2). \tag{18}$$

It is apparent that if $(D_1, D_2) < \mathcal{D}(X_1X_2)$, the ESLB of the joint, marginal and conditional rate distortion functions yield their corresponding rates, then the equations in (16)-(18) imply equalities in (10).

C. Successive Refinement

Successive refinement refers to a source coding scenario where a source is coded in multiple stages from coarser descriptions to finer descriptions and the description at each stage is optimal [26].

Consider a sequence of i.i.d random variables U^n where each U_i is drawn from a source alphabet $\mathcal{U}$ with distribution $p(u)$. Denote the reconstruction alphabet as $\hat{\mathcal{U}}$. Let $d : \mathcal{U} \times \hat{\mathcal{U}} \to [0, \infty)$ be the single letter distortion measure on $\mathcal{U} \times \hat{\mathcal{U}}$ which induces the distortion measure on $\mathcal{U}^n \times \hat{\mathcal{U}}^n$ according to

$$d(u^n, \hat{u}^n) = \frac{1}{n}\sum_{i=1}^{n} d(u_i, \hat{u}_i). \tag{19}$$

The U^n sequence is first described by a message at rate R_1, which incurs distortion D_1, then an addendum is provided to the first message at rate $(R_2 - R_1)$ so that the message now has distortion D_2 $(D_1 \geq D_2)$. If $R_1 = R_U(D_1)$, $R_2 = R_U(D_2)$, we say that the sequence is successively refinable.

Definition 2 [26]: A source U is said to be *successively refinable* from distortion D_1 to distortion D_2 for $D_1 \geq D_2$ if there exists a sequence of encoding function $f_1 : \mathcal{U}^n \to \{1, \cdots, 2^{nR_1}\}$ and $f_2 : \mathcal{U}^n \to \{1, \cdots, 2^{n(R_2-R_1)}\}$ and reconstruction functions $g_1 : \{1, \cdots, 2^{nR_1}\} \to \hat{\mathcal{U}}^n$ and $g_2 : \{1, \cdots, 2^{nR_1}\} \times \{1, \cdots, 2^{n(R_2-R_1)}\} \to \hat{\mathcal{U}}^n$ such that for $\hat{U}^n_1 = g_1(f_1(U^n))$ and $\hat{U}^n_2 = g_2(f_1(U^n), f_2(U^n))$,

$$R_1 = R_U(D_1), \quad \lim_{n\to\infty} \sup Ed(U^n, \hat{U}^n_1) \leq D_1, \tag{20}$$

$$R_2 = R_U(D_2), \quad \lim_{n\to\infty} \sup Ed(U^n, \hat{U}^n_2) \leq D_2. \tag{21}$$

The source U is said to be successively refinable if it is successively refinable from distortion D_1 to distortion D_2 for every $D_1 \geq D_2$.

It is shown in [26] that successive refinement from a coarse description $\hat{U}_1$ with distortion D_1 to a finer description $\hat{U}_2$ with distortion D_2 can be achieved if and only if the conditional distributions $p(\hat{u}_1|u)$ and $p(\hat{u}_2|u)$ which achieve $R_U(D_i) = I(U; \hat{U}_i), i = 1, 2$ are such that $U, \hat{U}_2, \hat{U}_1$ form a Markov chain.

Theorem 1 [26]: U is successively refinable from distortion D_1 to D_2 $(D_1 \geq D_2)$ if and only if the optimal random variables $\hat{U}_l$ achieving $(D_i, R_U(D_i)), i = 1, 2$ satisfy the Markov chain

$$U - \hat{U}_2 - \hat{U}_1. \tag{22}$$

A similar definition of successive refinement applies to vector sources with individual distortion constraints and a similar result to Theorem 1 can also be obtained for vector sources [29]. We state the Markovian characterization of successive refinement for a pair of random variable (U, V) here.

Theorem 2 [29]: The source (U, V) is successively refinable from (D^1_U, D^1_V) to (D^2_U, D^2_V) with $(D^1_U, D^1_V) \geq (D^2_U, D^2_V)$ if and only if the optimal random variables $(\hat{U}_i, \hat{V}_i)$ achieving $\left((D^i_U, D^i_V), R_{UV}(D^i_U, D^i_V)\right)$ for $i = 1, 2$ satisfy the Markov chain

$$(U, V) - (\hat{U}_2, \hat{V}_2) - (\hat{U}_1, \hat{V}_1), \tag{23}$$

where $(D^1_U, D^1_V) \geq (D^2_U, D^2_V)$ is a shorthand notation for $D^1_U \geq D^2_U$ and $D^1_V \geq D^2_V$.

III. The Lossy Source Coding Interpretation of Wyner's Common Information

The common information defined in (1) equally applies to that of continuous random variables. However, such definitions are only meaningful when they are associated with concrete operational interpretations. In this section, we develop a lossy source coding interpretation of Wyner's common information using the Gray-Wyner network.

A. Lossy Gray-Wyner Source Coding

In Section II-A, Wyner's first approach to explain the common information of discrete random variables is based on the lossless source coding theorem of the Gray-Wyner network. Let (R_0, R_1, R_2) be the rate triple of the three messages. The set of triples (R_0, R_1, R_2) satisfying

$$R_0 + R_1 + R_2 = H(X_1, X_2)$$

is referred to as the "Pangloss plane" as $H(X_1, X_2)$ is the minimum sum rate needed to recover (X_1, X_2) with joint decoding, hence provides a natural lower bound on the sum rate for the Gray-Wyner source coding problem.

Let $\mathcal{R}_1$ be the set of all achievable rate triples (R_0, R_1, R_2). For the rate triples that lie on the intersection of the region $\mathcal{R}_1$ and the Pangloss plane, a total rate of $H(X_1, X_2)$ can be split into three parts for (R_0, R_1, R_2) and the source X_1, X_2 can be recovered losslessly at the receivers. Wyner's first approach actually shows that $C(X_1, X_2)$ is the minimum R_0 that lies on the intersection of the achievable region $\mathcal{R}_1$ and the Pangloss plane.

Motivated by this relationship of Wyner's common information and the lossless rate region $\mathcal{R}_1$, we explore the connection of Wyner's common information and the lossy rate distortion region of Gray-Wyner network, which provides a new interpretation of Wyner's common information.

The rate distortion region of the Gray-Wyner network is defined as follows. Let $d_i(x_i, \hat{x}_i), i = 1, 2$ be a give nonnegative per-letter distortion function for X_i. Define $\Delta_i, i = 1, 2$ to

be the average distortion between the ith component sequence of the encoder input and the ith decoder output,

$$\Delta_i = E[d_i(X_i^n, \hat{X}_i^n)] = \frac{1}{n}\sum_{k=1}^{n} E[d_i(X_{ik}, \hat{X}_{ik})]. \quad (24)$$

An (n, M_0, M_1, M_2) code with an average distortion vector (Δ_1, Δ_2) is said to be an $(n, M_0, M_1, M_2, \Delta_1, \Delta_2)$ rate distortion code.

For any $(D_1, D_2) \in \mathbb{R}_+^2$, a rate triple (R_0, R_1, R_2) is said to be (D_1, D_2)-*achievable* if for arbitrary $\epsilon > 0$, there exists, for n sufficiently large, an $(n, M_0, M_1, M_2, \Delta_1, \Delta_2)$ code such that

$$M_i \le 2^{n(R_i+\epsilon)}, \quad (25)$$

$$\Delta_i \le D_i + \epsilon, \quad i = 0, 1, 2. \quad (26)$$

The rate distortion region $\mathcal{R}_2(D_1, D_2)$ is defined as the set of all (D_1, D_2)-achievable rate triples (R_0, R_1, R_2).

This definition of lossy source coding applied both to discrete and continuous random variables. A characterization of the region $\mathcal{R}_2(D_1, D_2)$ is given below.

Theorem 3 [17]: If the source (X_1, X_2) has the property that there exist $\hat{x}_1 \in \hat{\mathcal{X}}_1$ and $\hat{x}_2 \in \hat{\mathcal{X}}_2$ such that

$$Ed_i(X_i, \hat{x}_i) < \infty, \quad i = 1, 2, \quad (27)$$

then $\mathcal{R}_2(D_1, D_2)$ is the union of all rate tuples (R_0, R_1, R_2) that satisfy

$$R_0 \ge I(X_1, X_2; W), \quad (28)$$

$$R_i \ge R_{X_i|W}(D_i), \quad i = 1, 2, \quad (29)$$

for some $W \sim p(w|x_1, x_2)$.

Similar to the lossless case, we can define a number R_0 to be (D_1, D_2)-*achievable* as follows. For any $(D_1, D_2) \ge 0$, a number R_0 is said to be (D_1, D_2)-*achievable* if for any $\epsilon > 0$, there exists, for n sufficiently large, an $(n, M_0, M_1, M_2, \Delta_1, \Delta_2)$ code with

$$M_0 \le 2^{nR_0}, \quad (30)$$

$$\sum_{i=0}^{2} \frac{1}{n}\log M_i \le R_{X_1X_2}(D_1, D_2) + \epsilon, \quad (31)$$

$$\Delta_1 \le D_1 + \epsilon, \quad \Delta_2 \le D_2 + \epsilon. \quad (32)$$

Here (31) is also referred to as the Pangloss bound for the lossy source coding problem with the Gray-Wyner network. $C_3(D_1, D_2)$ is defined as the infimum of all R_0's that are (D_1, D_2)-achievable. Thus, $C_3(D_1, D_2)$ is the minimum common message rate for the Gray-Wyner network with sum rate $R_{X_1X_2}(D_1, D_2)$ while satisfying the distortion constraint. Since $R_0 = R_{X_1X_2}(D_1, D_2)$ is always (D_1, D_2)-achievable, it is obvious that

$$C_3(D_1, D_2) \le R_{X_1X_2}(D_1, D_2). \quad (33)$$

The following theorem gives a precise characterization of $C_3(D_1, D_2)$.

Theorem 4: If the source (X_1, X_2) has the property that there exist $\hat{x}_1 \in \hat{\mathcal{X}}_1$ and $\hat{x}_2 \in \hat{\mathcal{X}}_2$ such that

$$Ed_i(X_i, \hat{x}_i) < \infty, \quad i = 1, 2, \quad (34)$$

then

$$C_3(D_1, D_2) = \tilde{C}(D_1, D_2), \quad (35)$$

where $\tilde{C}(D_1, D_2)$ is the solution to the following optimization problem:

$$\inf I(X_1, X_2; W) \quad (36)$$

subject to

$$R_{X_1|W}(D_1) + R_{X_2|W}(D_2) + I(X_1, X_2; W) = R_{X_1X_2}(D_1, D_2).$$

Proof: See Appendix A. ∎

The authors in [16] gave an alternative characterization of $C_3(D_1, D_2)$. Define

$$C^*(D_1, D_2) = \inf I(X_1, X_2; W),$$

where the infimum is taken over all joint distributions for $X_1, X_2, X_1^*, X_2^*, W$ such that

$$X_1^* - W - X_2^*, \quad (37)$$

$$(X_1, X_2) - (X_1^*, X_2^*) - W, \quad (38)$$

where (X_1^*, X_2^*) achieves $R_{X_1X_2}(D_1, D_2)$. It was shown in [16] that $C_3(D_1, D_2) = C^*(D_1, D_2)$. This, combined with Theorem 4, establishes that

$$\tilde{C}(D_1, D_2) = C^*(D_1, D_2). \quad (39)$$

$\tilde{C}(D_1, D_2)$ is derived from the rate distortion region $\mathcal{R}_2(D_1, D_2)$ given in Theorem 3 while the authors in [16] chose to derive $C^*(D_1, D_2)$ from an alternative characterization of $\mathcal{R}_2(D_1, D_2)$ given in [22]. In Appendix B, we provide a direct proof of (39) for completeness. Also, as given in Appendix B, a necessary condition for the equality condition in the optimization problem (36) is

$$R_{X_1X_2|W}(D_1, D_2) = R_{X_1|W}(D_1) + R_{X_2|W}(D_2).$$

B. The Relation of $C_3(D_1, D_2)$ and the Common Information

Given our characterization of $C_3(D_1, D_2)$ in Theorem 4, we now establish its connection with $C(X_1, X_2)$ which leads to a new interpretation of Wyner's common information. We begin with the following two lemmas.

Lemma 3: Let W be the random variable that achieves the common information of X_1 and X_2. If

$$R_{X_1X_2|W}(D_1, D_2) + C(X_1, X_2) = R_{X_1X_2}(D_1, D_2),$$

then

$$\tilde{C}(D_1, D_2) \le C(X_1, X_2). \quad (40)$$

Lemma 3 is a direct consequence of Theorem 4 as the Markov chain $X_1 - W - X_2$ implies $R_{X_1X_2|W}(D_1, D_2) = R_{X_1|W}(D_1) + R_{X_2|W}(D_2)$. Thus, the equality constraint in (36) is satisfied. Inequality (40) follows as

$$\tilde{C}(D_1, D_2) \le I(X_1, X_2; W) = C(X_1, X_2).$$

The next lemma gives a sufficient condition under which $\tilde{C}(D_1, D_2) \ge C(X_1, X_2)$ is true.

Lemma 4: For any distortion pair $(D_1, D_2) \geq 0$, if the rate distortion function satisfies

$$R_{X_1X_2}(D_1, D_2) = R_{X_1}(D_1) + R_{X_2}(D_1) - I(X_1; X_2), \quad (41)$$

then we have

$$\tilde{C}(D_1, D_2) \geq C(X_1, X_2).$$

Proof: See Appendix C. ∎

Theorem 4, Lemmas 3 and 4, together with the relations of marginal, joint and conditional rate distortion functions described in Lemmas 1 and 2, allow us to determine a region such that $C_3(D_1, D_2)$ equals to the common information.

Theorem 5: Let (X_1, X_2) be a pair of random variable with distribution $p(x_1, x_2)$ and alphabet $\mathcal{X}_1 \times \mathcal{X}_2$, where $\mathcal{X}_1$ and $\mathcal{X}_2$ are arbitrary measurable spaces that can be discrete or continuous. Let W be any random variable achieving the common information of (X_1, X_2). Let the reproduction alphabets $\hat{\mathcal{X}}_1 = \mathcal{X}_1$, $\hat{\mathcal{X}}_2 = \mathcal{X}_2$ and two per-letter distortion measures $d_1(x_1, \hat{x}_1)$, $d_2(x_2, \hat{x}_2)$ satisfy

$$d_i(x_i, \hat{x}_i) > d_i(x_i, x_i) = 0, \quad x_i \neq \hat{x}_i, \; i = 1, 2. \quad (42)$$

If the following conditions are satisfied:

1) For any $x_1 \in \mathcal{X}_1$, $x_2 \in \mathcal{X}_2$ and $w \in \mathcal{W}$, $p(w|x_1x_2) > 0$,
2) There exists an $\hat{x}_1 \in \hat{\mathcal{X}}_1$, $\hat{x}_2 \in \hat{\mathcal{X}}_2$ such that

$$Ed_i(X_i, \hat{x}_i) < \infty, \quad i = 1, 2, \quad (43)$$

then there exists a strictly positive surface $\gamma \triangleq (\gamma_1, \gamma_2)$ such that, for $0 \leq (D_1, D_2) \leq \gamma$,

$$C_3(D_1, D_2) = C(X_1, X_2). \quad (44)$$

Proof: Because of the first condition in Theorem 5, applying Lemma 2 to the random variable triple (X_1, X_2, W) yields that there exists a strictly positive surface $\mathcal{D}(X_1X_2|W)$ such that for any $0 \leq (D_1, D_2) \leq \mathcal{D}(X_1X_2|W)$, $R_{X_1X_2|W}(D_1, D_2)$, $R_{X_1X_2}(D_1, D_2)$ equal their corresponding ESLB which satisfy that

$$I(X_1, X_2; W) + R^{(L)}_{X_1X_2|W}(D_1, D_2) = R^{(L)}_{X_1X_2}(D_1, D_2). \quad (45)$$

Thus for any $0 \leq (D_1, D_2) \leq \mathcal{D}(X_1X_2|W)$,

$$I(X_1, X_2; W) + R_{X_1X_2|W}(D_1, D_2) = R_{X_1X_2}(D_1, D_2). \quad (46)$$

Also by Lemma 2, there exists a strictly positive surface $\mathcal{D}(X_1X_2)$ such that for any $0 \leq (D_1, D_2) \leq \mathcal{D}(X_1X_2)$,

$$R_{X_1}(D_1) + R_{X_2}(D_2) - I(X_1; X_2) = R_{X_1X_2}(D_1, D_2). \quad (47)$$

Since $\mathcal{D}(X_1X_2|W) \leq \mathcal{D}(X_1X_2)$, let $\gamma = \mathcal{D}(X_1X_2|W)$, both equalities (46) and (47) hold for $0 \leq (D_1, D_2) \leq \gamma$. Therefore, the conditions in both Lemmas 3 and 4 are satisfied, we have $\tilde{C}(D_1, D_2) = C(X_1, X_2)$ for $0 \leq (D_1, D_2) \leq \gamma$. Thus the proof is completed by Theorem 4. ∎

Remarks:

1) Theorem 5 shows that under quite general conditions, Wyner's common information is precisely the smallest common message rate $C_3(D_1, D_2)$ of the Gray-Wyner network for a certain range of distortion constraints when the total rate is arbitrarily close to the rate distortion function with joint decoding. As the common information is only a function of the joint distribution, hence is a constant for a given $p(x_1, x_2)$, it is surprising that the smallest common rate $C_3(D_1, D_2)$ remains constant even if the distortion constraints vary, as long as they are within a specific distortion region.
2) While Theorem 5 establishes that $C_3(D_1, D_2) = C(X_1, X_2)$ for $(D_1, D_2) \leq \gamma$, it does not specify the value of the positive distortion surface γ. By the proof of Theorem 5, we know γ is exactly $\mathcal{D}(X_1X_2|W)$, the critical region of distortion where $R_{X_1X_2|W}(D_1, D_2)$ equals its corresponding ESLB $R^{(L)}_{X_1X_2|W}(D_1, D_2)$. Furthermore, let $\mathcal{D}^c \triangleq (D_1^c, D_2^c)$ be the two-dimensional distortion surface such that $R_{X_1X_2}(D_1^c, D_2^c) = C(X_1, X_2)$, then we must have

$$\gamma \leq \mathcal{D}^c.$$

This is because if $\gamma > \mathcal{D}^c$, then there exists (D_1, D_2) such that $\gamma \geq (D_1, D_2) > \mathcal{D}^c$ and $C_3(D_1, D_2) \leq R_{X_1X_2}(D_1, D_2) < R_{X_1X_2}(D_1^c, D_2^c) = C(X_1, X_2)$, which contradicts Theorem 5.

C. $C_3(D_1, D_2)$ for Successively Refinable Sources

From the second remark of Theorem 5, we know $\gamma \leq \mathcal{D}^c$. Now let us consider a particular point on the surface $\mathcal{D}^c$, denoted by (D_1^0, D_2^0) and defined below. We will show that under the assumption that such a point exists on $\mathcal{D}^c$ and the sources are successively refinable, then $C_3(D_1, D_2)$ equals the common information for any $(D_1, D_2) \leq (D_1^0, D_2^0)$.

Let W be the auxiliary random variable that achieves $C(X_1, X_2)$. Suppose there exists a distortion pair (D_1^0, D_2^0) satisfying, for $i = 1, 2$,

$$\begin{aligned} R_{X_i}(D_i^0) &= I(X_i; W), \\ D_i^0 &= \inf_{\hat{x}_i(w)} Ed_i(X_i, \hat{X}_i^0(W)), \end{aligned} \quad (48)$$

where $\hat{x}_1^0(w), \hat{x}_2^0(w)$ are deterministic functions. Under this assumption, we can show that $R_{X_1X_2}(D_1^0, D_2^0) = I(X_1, X_2; W)$, which means (D_1^0, D_2^0) is on the surface $\mathcal{D}^c$. The joint rate distortion function $R_{X_1X_2}(D_1^0, D_2^0)$ not only equals to the common information but also is achieved by the auxiliary random variable W. Furthermore, it is easy to establish

$$C_3(D_1^0, D_2^0) = C(X_1, X_2), \quad (49)$$

using Lemma 4 and the fact that $C_3(D_1^0, D_2^0) \leq R_{X_1X_2}(D_1^0, D_2^0)$. This means that in the Gray-Wyner network, with the total rate equal to $R_{X_1X_2}(D_1^0, D_2^0)$, the scheme to transmit the pair of sources (X_1^n, X_2^n) within distortion constraints (D_1^0, D_2^0) is to communicate W to the two receivers using the common channel.

Let us now decrease the distortion constraints from (D_1^0, D_2^0) to $(D_1, D_2) \leq (D_1^0, D_2^0)$. The question is whether the rate $C(X_1, X_2)$ is $(D_1, D_2)-$achieveble, i.e., if it is possible to transmit the sources (X_1^n, X_2^n) with smaller distortions (D_1, D_2) with the sum rate at $R_{X_1X_2}(D_1, D_2)$ while keeping the common rate at $C(X_1, X_2)$. In the following theorem, we identify a sufficient condition under

which $C_3(D_1, D_2) = C(X_1, X_2)$ for successively refinable sources. This sufficient condition ensures the optimality of a two-stage encoding scheme: first encode the common message with rate $C(X_1, X_2)$ and we can obtain a coarse distortion (D_1^0, D_2^0), then encode the two private messages with rates $R_{X_1|W}(D_1)$ and $R_{X_2|W}(D_2)$. The successive refinement assumption guarantees that the two-step approach achieves the distortion (D_1, D_2) and the sum rate does not exceed the total rate $R_{X_1X_2}(D_1, D_2)$.

Theorem 6: Assume the source (X_1, X_2) has the property that there exist $\hat{x}_1 \in \hat{\mathcal{X}}_1$ and $\hat{x}_2 \in \hat{\mathcal{X}}_2$ such that

$$Ed_i(X_i, \hat{x}_i) < \infty, \quad i = 1, 2.$$

Let W be the auxiliary variable that achieves $C(X_1, X_2)$ and (D_1^0, D_2^0) be a distortion pair satisfying (48). If the source (X_1, X_2) is successively refinable from (D_1^0, D_2^0) to (D_1, D_2) for $(D_1, D_2) \le (D_1^0, D_2^0)$, and X_i is successively refinable from D_i^0 to D_i for $D_i \le D_i^0$, $i = 1, 2$, then,

$$C_3(D_1, D_2) = C(X_1, X_2).$$

Proof: See Appendix D. ■

In the following section, we will consider two examples involving successively refinable sources: the binary random variables and bivariate Gaussian variables. For these two cases, we compute explicitly the function $C_3(D_1, D_2)$ and establish its connection with $C(X_1, X_2)$. The distortion pair (D_1^0, D_2^0) satisfying (48) is identified for both cases, thus Theorem 6 can be directly applied.

IV. Examples

A. Binary Random Variables

Let $S \sim \text{Bern}(\theta)$ for $0 \le \theta \le 1$, i.e., $S \in \{0, 1\}$ and $P(S = 1) = \theta$. Let X_i, $i = 1, \cdots, N$, be the output of a binary symmetric channel (BSC) with crossover probability a_1 ($0 \le a_1 \le \frac{1}{2}$) and with S as input. The BSC channels are independent of each other. Thus,

$$p(x_1, \cdots, x_N|s) = \prod_{i=1}^{N} p(x_i|s),$$

where

$$p(x_i|s) = \begin{cases} 1 - a_1, & \text{if } x_i = s, \\ a_1, & \text{otherwise,} \end{cases}$$

for $x_i \in \{0, 1\}$. Therefore, the joint distribution of $X_1, X_2, \cdots, X_N$ is

$$\begin{aligned} &p(x_1, x_2, \cdots, x_N) \\ &\quad = \sum_{s\in\{0,1\}} p(s) \prod_{i=1}^{N} p(x_i|s), \\ &\quad = \theta a_1^{t_N}(1 - a_1)^{N-t_N} + (1 - \theta)(1 - a_1)^{t_N} a_1^{N-t_N}, \end{aligned} \tag{50}$$

where $t_N = \sum_{i=1}^{N} x_i$.

For $N = 2$, the joint distribution of X_1, X_2 is given by the following probability matrix,

$$\begin{bmatrix} \theta(1 - a_1)^2 + (1 - \theta)a_1^2 & a_1(1 - a_1) \\ a_1(1 - a_1) & \theta a_1^2 + (1 - \theta)(1 - a_1)^2 \end{bmatrix}. \tag{51}$$

It has been shown by Witsenhausen [13] that the common information of X_1, X_2 is achieved with W being S. That is

$$C(X_1, X_2) = I(X_1, X_2; S) = H(X_1, X_2) - 2h(a_1),$$

where $h(\cdot)$ is the binary entropy function. When $\theta = \frac{1}{2}$, (X_1, X_2) is a Doubly Symmetric Binary Source (DSBS) whose common information was derived by Wyner [4] using a different approach.

The above result can be generalized to the common information for N variables , each of which is the channel output of a BSC with the common input S.

Proposition 1: Let $S \sim \text{Bern}(\theta)$ and let X_i, $i = 1, \cdots, N$, be the output of independent BSCs with common input S and crossover probability $0 \le a_1 \le 1/2$. Then for any $N \ge 2$, the common information for $X_1, \cdots, X_N$ is given as

$$C(X_1, \cdots, X_N) = I(X_1, \cdots, X_N; S). \tag{52}$$

Proof: That $C(X_1, \cdots, X_N) \le I(X_1, \cdots, X_N; S)$ follows from the definition of the common information for multiple random variables [31]. The inequality $C(X_1, \cdots, X_N) \ge I(X_1, \cdots, X_N; S)$ can be proved by contradiction. Suppose there exists a W such that

$$\begin{aligned} C(X_1, \cdots, X_N) &= I(X_1, \cdots, X_N; W) \\ &< I(X_1, \cdots, X_N; S), \end{aligned} \tag{53}$$

i.e., $C(X_1, \cdots, X_N)$ is achieved by W and it is strictly less than $I(X_1, \cdots, X_N; S)$. Since W induces conditional independence of $X_1, \cdots, X_N$, we have, from (53),

$$\sum_{i=1}^{N} H(X_i|W) > \sum_{i=1}^{N} H(X_i|S).$$

Thus, there must exist two random variables X_k, X_j, $k, j \in \{1, \cdots, N\}$ such that

$$H(X_k|W) + H(X_j|W) > H(X_k|S) + H(X_j|S).$$

Given that the sequence $\{X_1, \cdots, X_N\}$ is exchangeable [30], $p(x_k, x_j)$ has the same joint distribution as $p(x_1, x_2)$. Thus,

$$\begin{aligned} C(X_1, X_2) &= C(X_k, X_j) \\ &\le I(X_k, X_j; W) \\ &< I(X_k, X_j; S) \\ &= I(X_1, X_2; S). \end{aligned}$$

This, however, contradicts the fact that S achieves $C(X_1, X_2)$. Thus the proposition is proved. ■

We now characterize the minimum common rate $C_3(D_1, D_2)$ for a DSBS.

Proposition 2: Consider a DSBS (X_1, X_2) with distribution

$$p(x_1, x_2) = \begin{cases} \frac{1}{2}(1 - a_0), & \text{if } x_1 = x_2, \\ \frac{1}{2}a_0, & \text{otherwise,} \end{cases} \tag{54}$$

where, without loss of generality, $0 \le a_0 \le 1/2$. Let a_1 be such that $a_0 = 2a_1(1 - a_1), 0 \le a_1 \le 1/2$. With Hamming

distortion $d_1 = d_2 = d_H$, we have

$$C_3(D_1, D_2) = \begin{cases} C(X_1, X_2), & (D_1, D_2) \in \mathcal{E}_{10}, \\ R_{X_1X_2}(D_1, D_2), & (D_1, D_2) \in \mathcal{E}_2 \cup \mathcal{E}_3, \\ 0, & (D_1, D_2) \geq (\frac{1}{2}, \frac{1}{2}), \end{cases} \tag{55}$$

$$C(X_1, X_2) \leq C_3(D_1, D_2) \leq R_{X_1X_2}(D_1, D_2), \quad (D_1, D_2) \in \mathcal{E}_{11}, \tag{56}$$

where

$$\begin{aligned} \mathcal{E}_{10} &= \{(D_1, D_2) : 0 \leq D_i \leq a_1, i = 1, 2\}, \\ \mathcal{E}_{11} &= \mathcal{E}_{10}^c \cap \{(D_1, D_2) : D_1 + D_2 - 2D_1D_2 \leq a_0\}, \\ \mathcal{E}_2 &= \mathcal{E}_{10}^c \cap \mathcal{E}_{11}^c \cap \left\{(D_1, D_2) : \max\left\{\frac{D_1 - D_2}{1-2D_2}, \frac{D_2 - D_1}{1-2D_1}\right\} \leq a_0\right\}, \\ \mathcal{E}_3 &= \mathcal{E}_{10}^c \cap \mathcal{E}_{11}^c \cap \mathcal{E}_2^c \cap \left\{(D_1, D_2) : D_i \leq \frac{1}{2}, i = 1, 2\right\}. \end{aligned} \tag{57}$$

Proof: For $X_i \sim \text{Bern}(1/2), i = 1, 2$ with Hamming distortion, the rate distortion function is

$$R_{X_i}(D_i) = \begin{cases} 1 - h(D_i), & 0 \leq D_i \leq \frac{1}{2}, \\ 0, & D_i \geq \frac{1}{2}. \end{cases}$$

The joint rate distortion function of the DSBS (X_1, X_2) is given by [29]

$$R_{X_1X_2}(D_1, D_2) = \begin{cases} 1 + h(a_0) - h(D_1) - h(D_2), & (D_1, D_2) \in \mathcal{E}_1, \\ 1 - (1 - a_0)h\left(\frac{D_1+D_2-a_0}{2(1-a_0)}\right) - a_0 h\left(\frac{D_1-D_2+a_0}{2a_0}\right), & \\ & (D_1, D_2) \in \mathcal{E}_2, \\ 1 - h(\min\{D_1, D_2\}), & (D_1, D_2) \in \mathcal{E}_3. \end{cases} \tag{58}$$

where $\mathcal{E}_1 = \mathcal{E}_{10} \cup \mathcal{E}_{11}$ with $\mathcal{E}_{10}, \mathcal{E}_{11}, \mathcal{E}_2$ and $\mathcal{E}_3$ defined in (57). Therefore, for this DSBS, $R_{X_1}(D_1) + R_{X_2}(D_2) - I(X_1; X_2) = R_{X_1X_2}(D_1, D_2)$, for $(D_1, D_2) \in \mathcal{E}_1$. From Lemma 4, we have for $(D_1, D_2) \in \mathcal{E}_1$,

$$C_3(D_1, D_2) \geq C(X_1, X_2). \tag{59}$$

On the other hand, let S be the binary random variable that achieves the common information of X_1, X_2. That is $S \sim \text{Bern}(1/2)$ and $p(x_i|s) = 1 - a_1$ if $s = x_i$ for $i = 1, 2$. Then the conditional rate distortion function $R_{X_i|S}(D_i)$ is given by [18]

$$R_{X_i|S}(D_i) = \begin{cases} h(a_1) - h(D_i), & 0 \leq D_i \leq a_1, \\ 0, & D_i \geq a_1. \end{cases}$$

Therefore, $R_{X_1|S}(D_1) + R_{X_2|S}(D_2) + I(X_1, X_2; S) = R_{X_1X_2}(D_1, D_2)$ is satisfied for $(D_1, D_2) \in \mathcal{E}_{10}$. From Theorem 4, $C_3(D_1, D_2) \leq C(X_1, X_2)$ for $(D_1, D_2) \in \mathcal{E}_{10}$. Together with (59) and given that $\mathcal{E}_{10} \subset \mathcal{E}_1$, we have proved that for $(D_1, D_2) \in \mathcal{E}_{10}$,

$$C_3(D_1, D_2) = C(X_1, X_2).$$

For $(D_1, D_2) \in \mathcal{E}_2$, we only need to show that $C_3(D_1, D_2) \geq R_{X_1X_2}(D_1, D_2)$. It was shown in [29] that the backward test channel that achieves $R_{X_1X_2}(D_1, D_2)$ is given by

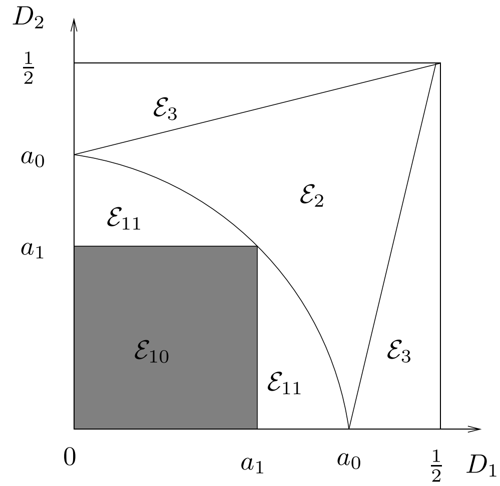

Fig. 3. The distortion regions $\mathcal{E}_{10}, \mathcal{E}_{11}, \mathcal{E}_2$ and $\mathcal{E}_3$ for the DSBS. $C_3(D_1, D_2) = C(X_1, X_2)$ in the shaded region.

$$\begin{aligned} X_1 &= \hat{X}_1 + Z_1, \\ X_2 &= \hat{X}_2 + Z_2, \end{aligned}$$

where both $\hat{X}_1, \hat{X}_2$ and Z_1, Z_2 are binary vectors independent of each other with the probability mass functions given respectively as

$$P_{\hat{X}_1\hat{X}_2} = \begin{bmatrix} \frac{1}{2} & 0 \\ 0 & \frac{1}{2} \end{bmatrix},$$

$$P_{Z_1Z_2} = \frac{1}{2}\begin{bmatrix} 2 - a_0 - D_1 - D_2 & D_2 - D_1 + a_0 \\ D_1 - D_2 + a_0 & D_1 + D_2 - a_0 \end{bmatrix}.$$

Therefore, $(\hat{X}_1, \hat{X}_2)$ that achieves $R_{X_1X_2}(D_1, D_2)$ satisfies

$$\hat{X}_2 = \hat{X}_1.$$

For the characterization $C^*(D_1, D_2)$ of $C_3(D_1, D_2)$, any W satisfying the Markov chain $\hat{X}_1 - W - \hat{X}_1$ must satisfy $H(\hat{X}_1|W) = 0$. Thus, $\hat{X}_1$ is a function of W and we have

$$\begin{aligned} I(X_1, X_2; W) &= I(X_1, X_2; W, \hat{X}_1) \\ &\geq I(X_1, X_2; \hat{X}_1) \\ &= R_{X_1X_2}(D_1, D_2). \end{aligned}$$

Therefore, $C_3(D_1, D_2) = R_{X_1X_2}(D_1, D_2)$.

The region $\mathcal{E}_3$ is a degenerated one. For example, $R_{X_1X_2}(D_1, D_2) = R_{X_1}(D_1)$ if $a_0 < \frac{D_2-D_1}{1-2D_1}$ and $D_i \leq \frac{1}{2}$, $i = 1, 2$. This implies that the optimal coding scheme is to ignore X_2 and optimally compress X_1. Then $\hat{X}_2$ can be estimated from $\hat{X}_1$ with distortion less than D_2. The case of $a_0 < \frac{D_1-D_2}{1-2D_2}$ is dealt with similarly. Hence, similar to the region $\mathcal{E}_2$, $C_3(D_1, D_2) = R_{X_1X_2}(D_1, D_2)$. ∎

The characterization of $C_3(D_1, D_2)$ is plotted in Fig. 3 as a function of the distortion constraints. $C_3(D_1, D_2) = C(X_1, X_2)$ in the shaded region. For the symmetric distortion constraint, $D_1 = D_2 = D$, the relation of $C_3(D, D)$ and D for the DSBS is given in Fig. 4.

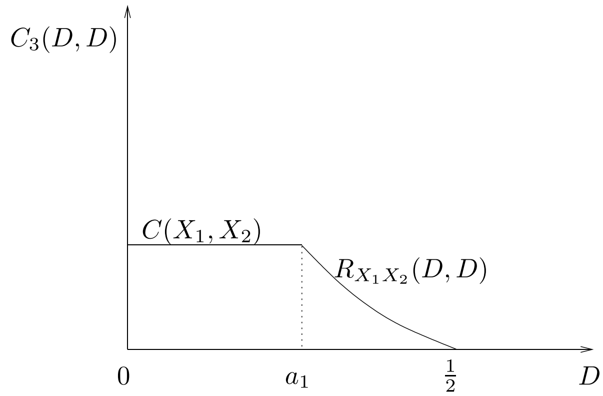

Fig. 4. The relation of $C_3(D, D)$ and D for the DSBS with $D_1 = D_2 = D$.

Remarks:

- The claim $C_3(D_1, D_2) = C(X_1, X_2)$ for $(D_1, D_2) \in \mathcal{E}_{10}$ can also be proved using Theorem 6. $R_{X_1X_2}(a_1, a_1)$ is achieved by the backward test channel $p_b(x_1, x_2|s) = p(x_1|s)p(x_2|s)$. The vector source (X_1, X_2) is successively refinable for any $(D_1, D_2) \le (a_1, a_1)$ [29] and the scalar source X_i is successively refinable for any $D_i \le a_1$, $i = 1, 2$ [26]. Thus by Theorem 6, $C_3(D_1, D_2) = C(X_1, X_2)$ for $(D_1, D_2) \le (a_1, a_1)$.
- We have the full characterization of $C_3(D_1, D_2)$ in the distortion region except the region $\mathcal{E}_{11}$. From the proof of Proposition 2, we know that $C_3(D_1, D_2) \ge C(X_1, X_2)$ for $(D_1, D_2) \in \mathcal{E}_{11}$, but the exact value of $C_3(D_1, D_2)$ in this region remains unknown.
- Let $(D_1, D_2) \le (D_1', D_2') \le (a_1, a_1)$, then the rate $R_{X_1X_2}(D_1', D_2')$ is $(D_1, D_2)-$achievable in the Gray-Wyner network, i.e., $R_{X_1X_2}(D_1', D_2') \ge C_3(D_1, D_2)$.
  To show this, let $(\hat{X}_1, \hat{X}_2)$ achieve $R_{X_1X_2}(D_1', D_2')$. The backward test channel that achieves $R_{X_1X_2}(D_1', D_2')$ satisfies $p_b(x_1, x_2|\hat{x}_1\hat{x}_2) = p_b(x_1|\hat{x}_1)p_b(x_2|\hat{x}_2)$ where
  $$p_b(x_i|\hat{x}_i) = \begin{cases} 1 - D_i', & \text{if } x_i = \hat{x}_i, \\ D_i', & \text{Otherwise.} \end{cases}$$
  for $i = 1, 2$. Then for $(D_1, D_2) \le (D_1', D_2') \le (a_1, a_1)$, let the rate allocation of R_0, R_1, R_2 in the Gray-Wyner network be
  $$\begin{aligned} R_0 &= R_{X_1X_2}(D_1', D_2') \\ &= 1 + h(a_0) - h(D_1') - h(D_2'), \\ R_i &= R_{X_i|\hat{X}_1\hat{X}_2}(D_i) = R_{X_i|\hat{X}_i}(D_i) \\ &= h(D_i') - h(D_i), i = 1, 2. \end{aligned} \tag{60}$$
  Since R_0, R_1 and R_2 in (60) sum up to $R_{X_1X_2}(D_1, D_2)$, $R_{X_1X_2}(D_1', D_2')$ is $(D_1, D_2)-$achievable.
  The minimal R_0 satisfying (60) is exactly $C(X_1, X_2)$, which is achieved by letting $(D_1', D_2') = (a_1, a_1)$.

B. Gaussian Random Variables

In this section we consider bivariate Gaussian random variables X_1, X_2 with zero mean and covariance matrix
$$K_2 = \begin{bmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{bmatrix}. \tag{61}$$
The common information between this pair of Gaussian random variables is given in the following theorem.

Theorem 7: For two joint Gaussian random variables X_1, X_2 with covariance matrix K_2, the common information is
$$C(X_1, X_2) = \frac{1}{2}\log\frac{1+\rho}{1-\rho}. \tag{62}$$

Proof: See Appendix E. ■

The above result generalizes to multi-variate Gaussian random variables satisfying a certain covariance matrix structure, the proof of which can be constructed in a similar fashion.

Theorem 8: For N joint Gaussian random variables $X_1, X_2, \cdots, X_N$ with covariance matrix K_N,
$$K_N = \begin{bmatrix} 1 & \rho & \cdots & \rho \\ \rho & 1 & \cdots & \rho \\ \cdot & \cdot & \cdots & \cdot \\ \rho & \rho & \cdots & 1 \end{bmatrix}, \tag{63}$$
the common information is
$$C(X_1, X_2, \cdots, X_N) = \frac{1}{2}\log\left(1 + \frac{N\rho}{1-\rho}\right). \tag{64}$$

We now characterize the minimum common rate $C_3(D_1, D_2)$ in the Gray-Wyner lossy source coding network for bivariate Gaussian random variables with covariance matrix K_2 in equation (61). It was shown in [16] that for symmetric distortion, i.e., $D_1 = D_2 = D$,
$$C_3(D, D) = \begin{cases} C(X_1, X_2), & 0 \le D \le 1-\rho, \\ R_{X_1X_2}(D, D), & 1-\rho \le D \le 1, \\ 0, & D \ge 1. \end{cases} \tag{65}$$
We characterize $C_3(D_1, D_2)$ for general distortion (D_1, D_2) in the following proposition.

Proposition 3: For bivariate Gaussian random variables X_1, X_2 with zero mean, covariance matrix K_2 and squared error distortion, we have that
$$C_3(D_1, D_2) = \begin{cases} C(X_1, X_2), & (D_1, D_2) \in \mathcal{D}_{10}, \\ R_{X_1X_2}(D_1, D_2), & (D_1, D_2) \in \mathcal{D}_2 \cup \mathcal{D}_3, \\ 0, & (D_1, D_2) \ge (1, 1), \end{cases} \tag{66}$$
$$C(X_1, X_2) \le C_3(D_1, D_2) \le R_{X_1X_2}(D_1, D_2), \quad (D_1, D_2) \in \mathcal{D}_{11}, \tag{67}$$
where
$$\begin{aligned} &\mathcal{D}_{10} = \{(D_1, D_2) : 0 \le D_i \le 1-\rho, i = 1, 2\}, \\ &\mathcal{D}_{11} = \mathcal{D}_{10}^c \cap \{(D_1, D_2) : D_1 + D_2 - D_1D_2 \le 1-\rho^2\}, \\ &\mathcal{D}_2 = \mathcal{D}_{10}^c \cap \mathcal{D}_{11}^c \cap \left\{(D_1, D_2) : \min\left\{\frac{1-D_1}{1-D_2}, \frac{1-D_2}{1-D_1}\right\} \ge \rho^2\right\}, \\ &\mathcal{D}_3 = \mathcal{D}_{10}^c \cap \mathcal{D}_{11}^c \cap \mathcal{D}_2^c \cap \{(D_1, D_2) : D_i \le 1, i = 1, 2\}. \end{aligned} \tag{68}$$

Proof: The joint rate distortion function for Gaussian random variables with squared error distortion [27]–[29] is given by

$$R_{X_1X_2}(D_1, D_2) = \begin{cases} \frac{1}{2}\log\frac{1-\rho^2}{D_1D_2}, & (D_1, D_2) \in \mathcal{D}_1, \\ \frac{1}{2}\log\frac{1-\rho^2}{D_1D_2-\left(\rho-\sqrt{(1-D_1)(1-D_2)}\right)^2}, & (D_1, D_2) \in \mathcal{D}_2, \\ \frac{1}{2}\log\frac{1}{\min\{D_1, D_2\}}, & (D_1, D_2) \in \mathcal{D}_3, \end{cases} \tag{69}$$

where $\mathcal{D}_1 = \mathcal{D}_{10} \cup \mathcal{D}_{11}$. The marginal rate distortion function for $X_i \sim \mathcal{N}(0, 1), i = 1, 2$, is

$$R_{X_i}(D_i) = \begin{cases} \frac{1}{2}\log\frac{1}{D_i}, & 0 \le D_i \le 1, \\ 0, & D_i \ge 1. \end{cases}$$

Therefore, $R_{X_1}(D_1) + R_{X_2}(D_2) - I(X_1; X_2) = R_{X_1X_2}(D_1, D_2)$, for $(D_1, D_2) \in \mathcal{D}_1$. From Lemma 4, for $(D_1, D_2) \in \mathcal{D}_1$,

$$C_3(D_1, D_2) \ge C(X_1, X_2).$$

On the other hand, the random variable W in the following decomposition of X_1 and X_2 achieves the common information

$$X_i = \sqrt{\rho}W + \sqrt{1-\rho}N_i, \quad i = 1, 2. \tag{70}$$

where W, N_1, N_2 are mutually independent standard Gaussian random variables. The conditional distribution of X given W is Gaussian distribution with variance $1-\rho$. Hence, for $i = 1, 2$, the conditional rate distortion function is

$$R_{X_i|W}(D_i) = \begin{cases} \frac{1}{2}\log\frac{1-\rho}{D_i}, & 0 \le D_i \le 1-\rho, \\ 0, & D_i \ge 1-\rho. \end{cases} \tag{71}$$

The condition $R_{X_1|W}(D_1) + R_{X_2|W}(D_2) + I(X_1, X_2; W) = R_{X_1X_2}(D_1, D_2)$ is satisfied for $(D_1, D_2) \in \mathcal{D}_{10}$. From Theorem 4, $C_3(D_1, D_2) \le C(X_1, X_2)$ for $(D_1, D_2) \in \mathcal{D}_{10}$. Since, $\mathcal{D}_{10} \subseteq \mathcal{D}_1$, we proved that for $(D_1, D_2) \in \mathcal{D}_{10}$,

$$C_3(D_1, D_2) = C(X_1, X_2).$$

For $(D_1, D_2) \in \mathcal{D}_2$, it was shown in [29] that $(\hat{X}_1, \hat{X}_2)$ that achieves $R_{X_1X_2}(D_1, D_2)$ satisfies

$$\hat{X}_2 = \sqrt{\frac{1-D_2}{1-D_1}}\hat{X}_1.$$

Hence, using the characterization $C^*(D_1, D_2)$, it is easy to show that the W satisfying the Markov chains (37) and (38) must satisfy two Markov chains

$$X_1X_2 - \hat{X}_1 - W - \hat{X}_2,$$
$$X_1X_2 - \hat{X}_2 - W - \hat{X}_1.$$

Therefore, we have

$$I(X_1, X_2; W) = I(X_1, X_2; \hat{X}_1) = I(X_1, X_2; \hat{X}_1, \hat{X}_2),$$

which proves $C_3(D_1, D_2) = R_{X_1X_2}(D_1, D_2)$.

The region $\mathcal{D}_3$ is a degenerated one. For example, $R_{X_1X_2}(D_1, D_2) = R_{X_1}(D_1)$ if $\frac{1-D_2}{1-D_1} < \rho^2$, this means that the correlation between X_1 and X_2 is so strong that the optimal coding scheme is to encode X_1 to within distortion D_1 and ignore X_2. Then $\hat{X}_2$ can be estimated from $\hat{X}_1$. We have

$$\hat{X}_2 = \rho\hat{X}_1.$$

The case of $\frac{1-D_1}{1-D_2} < \rho^2$ is dealt with similarly. Hence, we have $C_3(D_1, D_2) = R_{X_1X_2}(D_1, D_2)$. ■

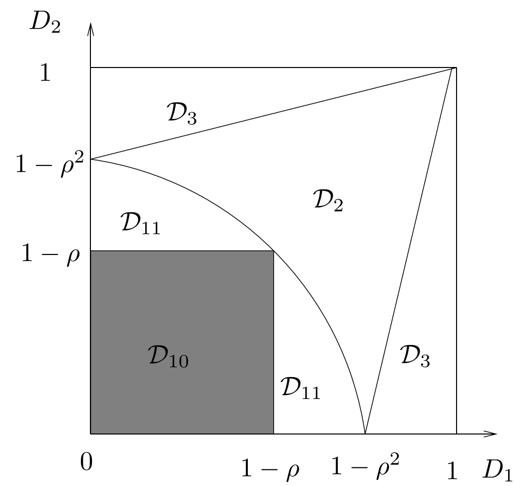

Fig. 5. The distortion regions $\mathcal{D}_{10}, \mathcal{D}_{11}, \mathcal{D}_2$ and $\mathcal{D}_3$ for bivariate Gaussian random variables. $C_3(D_1, D_2) = C(X_1, X_2)$ in the shaded region.

The characterization of $C_3(D_1, D_2)$ is plotted in Fig. 5 as a function of the distortion constraints. $C_3(D_1, D_2) = C(X_1, X_2)$ in the shaded region.

Remarks:

- Similar to the binary case, the claim $C_3(D_1, D_2) = C(X_1, X_2)$ for $(D_1, D_2) \in \mathcal{D}_{10}$ can also be proved using Theorem 6. This is because for the bivariate Gaussian random variables with covariance matrix K_2, $R_{X_1X_2}(1-\rho, 1-\rho)$ is achieved by the backward test channel $p_b(x_1, x_2|w) = p(x_1|w)p(x_2|w)$, (X_1, X_2) is successively refinable for any $(D_1, D_2) \le (1-\rho, 1-\rho)$ [29] and X_i is successively refinable for $D_i \le 1-\rho$, $i = 1, 2$ [26].
- Similarly, $C_3(D_1, D_2) \ge C(X_1, X_2)$ for $(D_1, D_2) \in \mathcal{D}_{11}$ but the exact characterization of $C_3(D_1, D_2)$ remains unknown in this region.
- Let $(D_1, D_2) \le (D_1', D_2') \le (1-\rho, 1-\rho)$, then the rate $R_{X_1X_2}(D_1', D_2')$ is $(D_1, D_2)-$achievable in the Gray-Wyner network, i.e., $R_{X_1X_2}(D_1', D_2') \ge C_3(D_1, D_2)$. This is because for $(D_1', D_2') \in \mathcal{E}_{10}$, the joint rate distortion function $R_{X_1X_2}(D_1', D_2')$ is achieved by Gaussian distributed $(\hat{X}_1, \hat{X}_2)$ satisfying $X_1 - \hat{X}_1 - \hat{X}_2 - X_2$ where the covariance matrix of $(\hat{X}_1, \hat{X}_2)$ is [29]

$$K_{\hat{X}_1\hat{X}_2} = \begin{bmatrix} 1-D_1' & \rho \\ \rho & 1-D_2' \end{bmatrix}.$$

Then for $(D_1, D_2) \le (D_1', D_2') \le (1-\rho, 1-\rho)$, let the rate allocation of R_0, R_1, R_2 for the Gray-Wyner network

be as follows:

$$R_0 = R_{X_1X_2}(D_1', D_2') = \frac{1}{2}\log\frac{1-\rho^2}{D_1'D_2'},$$
$$R_i = R_{X_i|\hat{X}_1\hat{X}_2}(D_i) = R_{X_i|\hat{X}_i}(D_i) = \frac{1}{2}\log\frac{D_i'}{D_i}, i = 1, 2. \tag{72}$$

R_0, R_1 and R_2 in (72) sum up to $R_{X_1X_2}(D_1, D_2)$, so $R_{X_1X_2}(D_1', D_2')$ is $(D_1, D_2)-$achievable.

Therefore, in the Gray-Wyner network, we can use the rate allocation in (72) to achieve the distortion $(D_1, D_2) \le (1-\rho, 1-\rho)$ for any $(D_1, D_2) \le (D_1', D_2') \le (1-\rho, 1-\rho)$. The minimal R_0 satisfying (72) is exactly $C(X_1, X_2)$, which is achieved by letting $(D_1', D_2') = (1-\rho, 1-\rho)$.

V. Conclusion

We have generalized the definition of Wyner's common information and expanded its practical significance by providing a new operational interpretation. We have derived a lossy source coding interpretation of Wyner's common information using the Gray-Wyner network. In particular, it is established that Wyner's common information is precisely the smallest common message rate when the total rate is arbitrarily close to the rate distortion function with joint decoding. A surprising observation is that such equality holds independent of the values of distortion constraints as long as the distortions are within some distortion region. Two examples, the doubly symmetric binary source under Hamming distortion and bivariate Gaussian source under squared-error distortion, are used to illustrate the lossy source coding interpretation of Wyner's common information. The common information for bivariate Gaussian source and its extension to the multi-variate case have also been computed explicitly.

While the lossy source coding interpretation of Wyner's common information presented in this paper is limited to $N = 2$ random variables, the results can be extended to arbitrary N random variables in a straightforward manner.

Appendix A
Proof of Theorem 4

We first show that $C_3(D_1, D_2) \ge \tilde{C}(D_1, D_2)$. Let R_0 be (D_1, D_2)-achievable, then for any $\epsilon > 0$, there exists an $(n, M_0, M_1, M_2, \Delta_1, \Delta_2)$ code such that

$$M_0 \le 2^{nR_0}, \tag{73}$$
$$\sum_{i=0}^{2}\frac{1}{n}\log M_i \le R_{X_1X_2}(D_1, D_2) + \epsilon, \tag{74}$$
$$\Delta_1 \le D_1 + \epsilon, \quad \Delta_2 \le D_2 + \epsilon. \tag{75}$$

Let $R_i' = \frac{1}{n}\log M_i$, for $i = 0, 1, 2$, then we know that (R_0', R_1', R_2') is (D_1, D_2)-achievable. From Theorem 3, there exists a W such that

$$R_0' \ge I(X_1, X_2; W), \tag{76}$$
$$R_i' \ge R_{X_i|W}(D_i), \quad i = 1, 2. \tag{77}$$

Therefore, for any $\epsilon > 0$, we have

$$\begin{aligned} &R_{X_1X_2}(D_1, D_2) + \epsilon \\ &\quad\ge \sum_{i=0}^{2} R_i', &(78)\\ &\quad\ge I(X_1, X_2; W) + \sum_{i=1}^{2} R_{X_i|W}(D_i), &(79)\\ &\quad\ge I(X_1, X_2; W) + R_{X_1X_2|W}(D_1, D_2), &(80)\\ &\quad\ge R_{X_1X_2}(D_1, D_2), &(81)\end{aligned}$$

where (78) is from the inequalities (74) and the definitions of $R_i', i = 0, 1, 2$, (79) is from (76) and (77), (80) is from (11b) and (81) comes from (10b).

Let $\epsilon \to 0$, then the left-hand side (LHS) and right-hand side (RHS) of the above inequalities become the same, all the inequalities must be equalities. Thus, we have

$$\begin{aligned} &I(X_1, X_2; W) + R_{X_1|W}(D_1) + R_{X_2|W}(D_2) \\ &\quad = R_{X_1X_2}(D_1, D_2). \end{aligned} \tag{82}$$

Hence, if R_0 is (D_1, D_2)-achievable, there exists a W such that $R_0 \ge I(X_1, X_2; W)$ and (82) is true. It shows that $C_3(D_1, D_2) \ge \tilde{C}(D_1, D_2)$.

Next we show $C_3(D_1, D_2) \le \tilde{C}(D_1, D_2)$. Let W' be any random variable satisfying the equality condition in the optimization problem (36). For any $R_0 > I(X_1, X_2; W')$ and $\epsilon > 0$, let

$$\epsilon_1 = \min\left\{\frac{\epsilon}{3}, R_0 - I(X_1, X_2; W')\right\}, \tag{83}$$

and hence $\epsilon_1 > 0$.

From Theorem 3, since the rate triple $(I(X_1, X_2; W'), R_{X_1|W'}(D_1), R_{X_2|W'}(D_2))$ is (D_1, D_2)-achievable, there exists an $(n, M_0, M_1, M_2, \Delta_1, \Delta_2)$ code such that

$$\frac{1}{n}\log M_0 \le I(X_1, X_2; W') + \epsilon_1 \le R_0, \tag{84}$$
$$\frac{1}{n}\log M_i \le R_{X_i|W'}(D_i) + \epsilon_1, \quad i = 1, 2. \tag{85}$$

Sum over (84) and (85), we get

$$\begin{aligned} \sum_{i=0}^{2}\frac{1}{n}\log M_i &\le I(X_1, X_2; W') + \sum_{i=1}^{2} R_{X_i|W'}(D_i) + 3\epsilon_1 \\ &= R_{X_1X_2}(D_1, D_2) + 3\epsilon_1, &(86)\\ &\le R_{X_1X_2}(D_1, D_2) + \epsilon, &(87)\end{aligned}$$

where (86) follows from the fact that W' satisfies the equality condition in the optimization problem (36) and inequality (87) is from (83).

This proves that R_0 is (D_1, D_2)-achievable, thus completes the proof of $C_3(D_1, D_2) \le \tilde{C}(D_1, D_2)$.

Appendix B
Direct Proof of $\tilde{C}(D_1, D_2) = C^*(D_1, D_2)$

First we show that $\tilde{C}(D_1, D_2) \ge C^*(D_1, D_2)$.

Let W be any random variable satisfying the equality condition in the optimization problem (36) and let $\hat{X}_1, \hat{X}_2$ be random variables that achieve $R_{X_1|W}(D_1)$ and $R_{X_2|W}(D_2)$, i.e.,

$$R_{X_1X_2}(D_1, D_2) = I(X_1, X_2; W) + R_{X_1|W}(D_1) + R_{X_2|W}(D_2), \tag{88}$$

$$R_{X_1|W}(D_1) = I(X_1; \hat{X}_1|W), \tag{89}$$

$$R_{X_2|W}(D_2) = I(X_2; \hat{X}_2|W), \tag{90}$$

$$E[d_1(X_1, \hat{X}_1)] \le D_1, \tag{91}$$

$$E[d_2(X_2, \hat{X}_2)] \le D_2. \tag{92}$$

Without loss of generality, we can assume that the joint distribution of $(X_1, X_2, \hat{X}_1, \hat{X}_2, W)$ factors as

$$p(x_1, x_2, \hat{x}_1, \hat{x}_2, w) = p(x_1, x_2, w)p(\hat{x}_1|x_1, w)p(\hat{x}_2|x_2, w),$$

because the distortion D_1 is independent of X_2 and D_2 is independent of X_1. To establish

$$R_{X_1X_2|W}(D_1, D_2) = R_{X_1|W}(D_1) + R_{X_2|W}(D_2), \tag{93}$$

we combine (88) and the inequalities below

$$R_{X_1X_2|W}(D_1, D_2) + I(X_1, X_2; W) \ge R_{X_1X_2}(D_1, D_2),$$
$$R_{X_1X_2|W}(D_1, D_2) \le R_{X_1|W}(D_1) + R_{X_2|W}(D_2),$$

from Lemma 1.

Therefore, together with (88)-(92), we have

$$\begin{aligned}
&R_{X_1X_2|W}(D_1, D_2)\\
&\ge I(X_1; \hat{X}_1|W) + I(X_2; \hat{X}_2|W)\\
&= H(\hat{X}_1|W) + H(\hat{X}_2|W) - H(\hat{X}_1|X_1, W) - H(\hat{X}_2|X_2, W)\\
&\ge H(\hat{X}_1, \hat{X}_2|W) - H(\hat{X}_1|X_1, W) - H(\hat{X}_2|X_2, W)\\
&= H(\hat{X}_1, \hat{X}_2|W) - H(\hat{X}_1|W, X_1, X_2) - H(\hat{X}_2|W, X_1, X_2)\\
&= I(X_1, X_2; \hat{X}_1, \hat{X}_2|W)\\
&\ge R_{X_1X_2|W}(D_1, D_2).
\end{aligned}$$

As the LHS and RHS of the above inequalities are the same, all the inequalities must be equalities so we have

$$I(\hat{X}_1; \hat{X}_2|W) = 0.$$

Furthermore we have

$$\begin{aligned}
&R_{X_1X_2}(D_1, D_2)\\
&\quad = I(X_1, X_2; W) + I(X_1; \hat{X}_1|W) + I(X_2; \hat{X}_2|W)\\
&\quad = I(X_1, X_2; W, \hat{X}_1, \hat{X}_2) - I(X_1, X_2; \hat{X}_1, \hat{X}_2|W)\\
&\qquad + I(X_1; \hat{X}_1|W) + I(X_2; \hat{X}_2|W)\\
&\quad = I(X_1, X_2; \hat{X}_1, \hat{X}_2) + I(X_1, X_2; W|\hat{X}_1, \hat{X}_2)\\
&\quad \ge I(X_1, X_2; \hat{X}_1, \hat{X}_2)\\
&\quad \ge R_{X_1X_2}(D_1, D_2).
\end{aligned}$$

The LHS and RHS of the above inequalities are the same, all the inequalities must be equalities so we have

$$I(X_1, X_2; W|\hat{X}_1, \hat{X}_2) = 0,$$
$$I(X_1, X_2; \hat{X}_1, \hat{X}_2) = R_{X_1X_2}(D_1, D_2).$$

Therefore, $X_1, X_2, \hat{X}_1, \hat{X}_2, W$ satisfy the Markov chains in (37) and (38) and $\hat{X}_1, \hat{X}_2$ achieve $R_{X_1X_2}(D_1, D_2)$. Thus, $\tilde{C}(D_1, D_2) \ge C^*(D_1, D_2)$.

Next we show that $\tilde{C}(D_1, D_2) \le C^*(D_1, D_2)$.

Let $X_1, X_2, X_1^*, X_2^*, W$ achieve $C^*(D_1, D_2)$. Therefore, they satisfy the Markov chains in (37) and (38) and $I(X_1, X_2; X_1^*, X_2^*) = R_{X_1X_2}(D_1, D_2)$ and $E[d_1(X_1, X_1^*)] \le D_1$, $E[d_2(X_2, X_2^*)] \le D_2$.

$$\begin{aligned}
&R_{X_1X_2}(D_1, D_2)\\
&= I(X_1, X_2; X_1^*, X_2^*)\\
&= I(X_1, X_2; W, X_1^*, X_2^*) &(94)\\
&= I(X_1, X_2; W) + I(X_1, X_2; X_1^*, X_2^*|W)\\
&= I(X_1, X_2; W) + H(X_1^*|W) + H(X_2^*|W) &(95)\\
&\quad - H(X_1^*, X_2^*|X_1, X_2, W) &(96)\\
&= I(X_1, X_2; W) + I(X_1; X_1^*|W) + I(X_2; X_2^*|W)\\
&\quad + H(X_1^*|X_1, W) + H(X_2^*|X_2, W) - H(X_1^*, X_2^*|X_1, X_2, W)\\
&\ge I(X_1, X_2; W) + I(X_1; X_1^*|W) + I(X_2; X_2^*|W)\\
&\quad + H(X_1^*|X_1, X_2, W) + H(X_2^*|X_1, X_2, W)\\
&\quad - H(X_1^*, X_2^*|X_1, X_2, W) &(97)\\
&= I(X_1, X_2; W) + I(X_1; X_1^*|W) + I(X_2; X_2^*|W)\\
&\quad + I(X_1^*; X_2^*|X_1, X_2, W)\\
&\ge I(X_1, X_2; W) + I(X_1; X_1^*|W) + I(X_2; X_2^*|W)\\
&\ge I(X_1, X_2; W) + R_{X_1|W}(D_1) + R_{X_2|W}(D_2)\\
&\ge I(X_1, X_2; W) + R_{X_1X_2|W}(D_1, D_2) &(98)\\
&\ge R_{X_1X_2}(D_1, D_2), &(99)
\end{aligned}$$

where (94) is from the Markov chain $(X_1, X_2) - (X_1^*, X_2^*) - W$, (96) is from the Markov chain $X_1^* - W - X_2^*$, (97) is because conditioning reduces entropy, (98) and (99) are by the properties of rate distortion functions. As the LHS and RHS of the above inequalities are the same, all the inequalities must be equalities so we have

$$I(X_1, X_2; W) + R_{X_1|W}(D_1) + R_{X_2|W}(D_2) = R_{X_1X_2}(D_1, D_2).$$

Therefore, $C^*(D_1, D_2) = I(X_1, X_2; W) \ge \tilde{C}(D_1, D_2)$.

Appendix C
Proof of Lemma 4

Let W be any random variable satisfying the equality condition in the optimization problem (36), that is

$$R_{X_1|W}(D_1) + R_{X_2|W}(D_2) + I(X_1, X_2; W) = R_{X_1X_2}(D_1, D_2). \tag{100}$$

Combined with (41), we have that

$$\begin{aligned}
&R_{X_1}(D_1) + R_{X_2}(D_2) - I(X_1; X_2)\\
&\quad = R_{X_1|W}(D_1) + R_{X_2|W}(D_2) + I(X_1, X_2; W) &(101)\\
&\quad \ge R_{X_1}(D_1) - I(X_1; W) + R_{X_2}(D_2) - I(X_2; W)\\
&\qquad + I(X_1, X_2; W) &(102)\\
&\quad = R_{X_1}(D_1) + R_{X_2}(D_2) - I(X_1; X_2) + I(X_1; X_2|W) &(103)\\
&\quad \ge R_{X_1}(D_1) + R_{X_2}(D_2) - I(X_1; X_2), &(104)
\end{aligned}$$

where equation (101) is from equations (100) and (41), inequality (102) comes from Lemma 1, (103) is by the chain rule and inequality (104) is by the fact that $I(X_1; X_2|W) \ge 0$.

Because the LHS of (101) is the same as the RHS of (104), we can conclude that all the inequalities above should be equalities. This implies $I(X_1; X_2|W) = 0$. Therefore, $\tilde{C}(D_1, D_2) \geq C(X_1, X_2)$ completes the proof.

Appendix D
Proof of Theorem 6

From Theorem 4, we only need to prove $\tilde{C}(D_1, D_2) = C(X_1, X_2)$.

First we show that for any $(D_1, D_2) \leq (D_1^0, D_2^0)$,

$$R_{X_1X_2|W}(D_1, D_2) + I(X_1, X_2; W) = R_{X_1X_2}(D_1, D_2). \quad (105)$$

We have the following inequality

$$\begin{aligned} R_{X_1X_2}(D_1^0, D_2^0) &\geq R_{X_1}(D_1^0) + R_{X_2}(D_2^0) - I(X_1; X_2) \quad (106)\\ &= I(X_1, X_2; W), \quad (107)\end{aligned}$$

where (106) is from (10c) and the equality (107) is from the definition of (D_1^0, D_2^0) in (48), the Markov chain $X_1 - W - X_2$, and the chain rule.

On the other hand,

$$R_{X_1X_2}(D_1^0, D_2^0) \leq I(X_1, X_2; \hat{X}_1^0, \hat{X}_2^0) \leq I(X_1, X_2; W), \quad (108)$$

where the first inequality is from the definition of rate distortion function and the second inequality is from the Markov chain $(X_1, X_2) - W - (\hat{X}_1^0, \hat{X}_2^0)$ and the chain rule. Combining (107) and (108), we have $R_{X_1X_2}(D_1^0, D_2^0) = I(X_1, X_2; \hat{X}_1^0, \hat{X}_2^0) = I(X_1, X_2; W)$.

Let $(\hat{X}_1, \hat{X}_2)$ be the random variables achieving $R_{X_1X_2}(D_1, D_2)$. As the vector source (X_1, X_2) is successively refinable under individual distortion constraints, by Theorem 2, we have the Markov chain $(X_1, X_2) - (\hat{X}_1, \hat{X}_2) - (\hat{X}_1^0, \hat{X}_2^0)$. Therefore,

$$\begin{aligned} &R_{X_1X_2}(D_1, D_2) - I(X_1, X_2; W)\\ &= I(X_1, X_2; \hat{X}_1, \hat{X}_2) - I(X_1, X_2; \hat{X}_1^0, \hat{X}_2^0)\\ &= I(X_1, X_2; \hat{X}_1, \hat{X}_2|\hat{X}_1^0, \hat{X}_2^0)\\ &\geq R_{X_1X_2|\hat{X}_1^0\hat{X}_2^0}(D_1, D_2)\\ &\geq R_{X_1X_2|W}(D_1, D_2),\end{aligned}$$

where the last inequality is from the Markov chain $(X_1, X_2) - W - (\hat{X}_1^0, \hat{X}_2^0)$. On the other hand, by Lemma 1, we have

$$R_{X_1X_2|W}(D_1, D_2) + I(X_1, X_2; W) \geq R_{X_1X_2}(D_1, D_2).$$

This establishes (105). Thus, from Lemma 3, $\tilde{C}(D_1, D_2) \leq C(X_1, X_2)$.

To complete the proof, we need to show

$$R_{X_1}(D_1) + R_{X_2}(D_2) - I(X_1; X_2) = R_{X_1X_2}(D_1, D_2), \quad (109)$$

which yields $\tilde{C}(D_1, D_2) \geq C(X_1, X_2)$ in view of Lemma 4.

From Lemma 1,

$$R_{X_1}(D_1) + R_{X_2}(D_2) - I(X_1; X_2) \leq R_{X_1X_2}(D_1, D_2).$$

Therefore, we only need to establish the other direction. For $i = 1, 2$, let $\hat{X}_i$ be the random variable achieving $R_{X_i}(D_i)$, then by the Markov property of successively refinable scalar source given in Theorem 1, we have the Markov chain $X_i - \hat{X}_i - \hat{X}_i^0$ for $D_i \leq D_i^0$. Therefore,

$$\begin{aligned} R_{X_i}(D_i) - I(X_i; W) &= I(X_i; \hat{X}_i) - I(X_i; \hat{X}_i^0)\\ &= I(X_i; \hat{X}_i|\hat{X}_i^0)\\ &\geq R_{X_i|\hat{X}_i^0}(D_i)\\ &\geq R_{X_i|W}(D_i), \quad (110)\end{aligned}$$

where (110) is from the Markov chain $X_i - W - \hat{X}_i^0$.

Using (110), we have

$$\begin{aligned} &R_{X_1}(D_1) + R_{X_2}(D_2) - I(X_1; X_2)\\ &\geq R_{X_1|W}(D_1) + I(X_1; W) + R_{X_2|W}(D_1)\\ &\quad + I(X_2; W) - I(X_1; X_2)\\ &= R_{X_1|W}(D_1) + R_{X_2|W}(D_2) + I(X_1, X_2; W)\\ &= R_{X_1X_2|W}(D_1, D_2) + I(X_1, X_2; W) \quad (111)\\ &= R_{X_1X_2}(D_1, D_2), \quad (112)\end{aligned}$$

where (111) is because $X_1 - W - X_2$ and the equation (11b), (112) is from the equation (105). This completes the proof.

Appendix E
Proof of Theorem 7

First, we will show that the common information of X_1, X_2 is only a function of the correlation coefficient ρ. To show this, let $\tilde{X}_i = \frac{1}{\sigma_i}X_i$, $i = 1, 2$, thus $\tilde{X}_1, \tilde{X}_2$ are jointly Gaussian distributed with zero mean and covariance matrix

$$\begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}. \quad (113)$$

We have the Markov chain that $\tilde{X}_1 - X_1 - X_2 - \tilde{X}_2$ and by the data processing inequality for Wyner's common information [13], $C(\tilde{X}_1, \tilde{X}_2) \leq C(X_1, X_2)$. On the other hand, we have the Markov chain that $X_1 - \tilde{X}_1 - \tilde{X}_2 - X_2$ and $C(\tilde{X}_1, \tilde{X}_2) \geq C(X_1, X_2)$. Thus, $C(\tilde{X}_1, \tilde{X}_2) = C(X_1, X_2)$. Without loss of generality, we will consider $\sigma_1^2 = \sigma_2^2 = 1$, i.e.,the covariance matrix is in the form (113) instead of (61).

Let

$$X_i = \sqrt{\rho}W + \sqrt{1-\rho}N_i, \quad i = 1, 2, \quad (114)$$

where W, N_1, N_2 are mutually independent standard Gaussian random variables. It is clear that X_1, X_2 are bivariate Gaussian with correlation coefficient ρ,

$$C(X_1, X_2) \leq I(X_1, X_2; W) = \frac{1}{2}\log\frac{1+\rho}{1-\rho}.$$

Next we will show that

$$C(X_1, X_2) \geq \frac{1}{2}\log\frac{1+\rho}{1-\rho}.$$

For any U that satisfies the Markov chain $X_1 - U - X_2$, let D_1 be the minimum mean square error (MMSE) of estimating X_1 using U, thus, $D_1 = E(X_1 - E(X_1|U))^2$. Similarly, let

$D_2 = E(X_2 - E(X_2|U))^2$. We now show that $I(X_1, X_2; U) \geq \frac{1}{2}\log\frac{1+\rho}{1-\rho}$.

$$\begin{aligned}
&I(X_1, X_2; U)\\
&\quad = H(X_1, X_2) - H(X_1|U) - H(X_2|U)\\
&\quad = I(X_1; U) + I(X_2; U) - I(X_1; X_2) && (115)\\
&\quad \geq I(X_1; E(X_1|U)) + I(X_2; E(X_2|U)) - I(X_1; X_2) && (116)\\
&\quad \geq R_{X_1}(D_1) + R_{X_2}(D_2) - I(X_1; X_2) && (117)\\
&\quad = \frac{1}{2}\log\frac{1-\rho^2}{D_1 D_2}, \text{ for } D_1 \leq 1, D_2 \leq 1,
\end{aligned}$$

where (115) is from the chain rule, (116) is from the Markov chains $X_1 - U - E(X_1|U)$, $X_2 - U - E(X_2|U)$ and (117) is by the definition of rate distortion function.

Next we show that $D_1 + D_2 \leq 2(1-\rho)$, $D_1 \leq 1$, $D_2 \leq 1$.

$$\begin{aligned}
&2(1-\rho)\\
&= E(X_1 - X_2)^2\\
&= E[X_1 - E(X_1|U) + E(X_1|U) - X_2]^2\\
&= E[X_1 - E(X_1|U)]^2 + E[E(X_1|U) - X_2]^2\\
&\quad + 2E[(X_1 - E(X_1|U))(E(X_1|U) - X_2)]\\
&= E[X_1 - E(X_1|U)]^2 + E[E(X_1|U) - X_2]^2 && (118)\\
&= E[X_1 - E(X_1|U)]^2 + E[E(X_1|U) - E(X_2|U)\\
&\quad + E(X_2|U) - X_2]^2\\
&= E[X_1 - E(X_1|U)]^2 + E[X_2 - E(X_2|U)]^2 + E[E(X_2|U)\\
&\quad - E(X_1|U))^2 + E[(X_2 - E(X_2|U))(E(X_2|U) - E(X_1|U))]\\
&= E[X_1 - E(X_1|U)]^2 + E[X_2 - E(X_2|U)]^2\\
&\quad + E[E(X_2|U) - E(X_1|U)]^2 && (119)\\
&\geq D_1 + D_2
\end{aligned}$$

where (118) is from

$$\begin{aligned}
&E[(X_1 - E(X_1|U))(E(X_1|U) - X_2)]\\
&\quad = E[(X_1 - E(X_1|U))E(X_1|U)] - E[(X_1 - E(X_1|U))X_2]\\
&\quad = -E[(X_1 - E(X_1|U))X_2]\\
&\quad = -E_{UX_2}[X_2 E_{X_1|U}[X_1 - E(X_1|U)]]\\
&\quad = -E_{UX_2}[X_2(E(X_1|U) - E(X_1|U))]\\
&\quad = 0,
\end{aligned}$$

and (119) is from

$$\begin{aligned}
&E[(X_2 - E(X_2|U))(E(X_2|U) - E(X_1|U))]\\
&\quad = E[(X_2 - E(X_2|U))E(X_2|U)]\\
&\qquad - E[(X_2 - E(X_2|U))E(X_1|U)]\\
&\quad = 0.
\end{aligned}$$

In addition, we have $D_1 = E[X_1 - E(X_1|U)]^2 = EX_1^2 - E[E(X_1|U)^2] \leq EX_1^2 = 1$. Thus,

$$\begin{aligned}
I(X_1, X_2; U) &\geq \frac{1}{2}\log\frac{1-\rho^2}{D_1 D_2}\\
&\geq \frac{1}{2}\log\frac{1-\rho^2}{\left(\frac{D_1+D_2}{2}\right)^2}\\
&\geq \frac{1}{2}\log\frac{1-\rho^2}{(1-\rho)^2}\\
&= \frac{1}{2}\log\frac{1+\rho}{1-\rho}.
\end{aligned}$$

Acknowledgment

The authors gratefully acknowledges Professor Paul Cuff of Princeton University for pointing out an error in an earlier proof of Theorem 7 given in [32].

References

[1] C. E. Shannon, "A mathematical theory of communication," *Bell Syst. Tech. J.*, vol. 27, no. 3, pp. 379–423, 1948.

[2] P. Gács and J. Körner, "Common information is much less than mutual information," *Problems Control Inf. Theory*, vol. 2, no. 2, pp. 149–162, 1973.

[3] R. Ahlswede and J. Körner, "On common information and related characteristics of correlated information sources," Presented at the 7th Prague Conf. Inf. Theory, 1974.

[4] A. D. Wyner, "The common information of two dependent random variables," *IEEE Trans. Inf. Theory*, vol. 21, no. 2, pp. 163–179, Mar. 1975.

[5] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. New York, NY, USA: Wiley, 1991.

[6] I. Csiszár and J. Körner, *Information Theory: Coding Theorems for Discrete Memoryless Systems*. New York, NY, USA: Academic, 1981.

[7] R. Ahlswede and I. Csiszár, "Common randomness in information theory and cryptography. I. Secret sharing," *IEEE Trans. Inf. Theory*, vol. 39, no. 4, pp. 1121–1132, Jul. 1993.

[8] R. Ahlswede and I. Csiszár, "Common randomness in information theory and cryptography. II. CR capacity," *IEEE Trans. Inf. Theory*, vol. 44, no. 1, pp. 225–240, Jan. 1998.

[9] U. M. Maurer, "Secret key agreement by public discussion from common information," *IEEE Trans. Inf. Theory*, vol. 39, no. 3, pp. 733–742, May 1993.

[10] S. Kamath and V. Anantharam, "A new dual to the Gács–Körner common information defined via the Gray–Wyner system," in *Proc. 48th Annu. Allerton Conf. Commun., Control Comput.*, Monticello, IL, USA, Sep. 2010, pp. 1340–1346.

[11] H. S. Witsenhausen, "On sequences of pairs of dependent random variables," *SIAM J. Appl. Math.*, vol. 28, no. 1, pp. 100–113, Jan. 1975.

[12] P. W. Cuff, H. H. Permuter, and T. M. Cover, "Coordination capacity," *IEEE Trans. Inf. Theory*, vol. 56, no. 9, pp. 4181–4206, Sep. 2010.

[13] H. S. Witsenhausen, "Values and bounds for the common information of two discrete random variables," *SIAM J. Appl. Math.*, vol. 31, no. 2, pp. 313–333, 1976.

[14] R. W. Yeung, *Information Theory and Network Coding*. New York, NY, USA: Springer, 2008.

[15] H. Tyagi, P. Narayan, and P. Gupta, "When is a function securely computable?" *IEEE Trans. Inf. Theory*, vol. 57, no. 10, pp. 6337–6350, Oct. 2011.

[16] K. Viswanatha, E. Akyol, and K. Rose, "Lossy common information of two dependent random variables," in *Proc. IEEE Symp. Inf. Thoery*, Cambridge, MA, USA, Jul. 2012, pp. 528–532.

[17] R. M. Gray and A. D. Wyner, "Source coding for a simple network," *Bell Syst. Tech. J.*, vol. 58, no. 9, pp. 1681–1721, Nov. 1974.

[18] R. M. Gray, "A new class of lower bounds to information rates of stationary sources via conditional rate-distortion functions," *IEEE Trans. Inf. Theory*, vol. 19, no. 4, pp. 480–489, Jul. 1973.

[19] B. M. Leiner, "An alternate proof of the composite lower bounds," *Inf. Control*, vol. 33, no. 1, pp. 72–86, Jan. 1977.

[20] R. M. Gray, “Conditional rate-distortion theory,” Stanford Electron. Lab., Stanford, CA, USA, Tech. Rep. 6502-2, Oct. 1972.

[21] G. Xu and B. Chen, “Information for inference,” *Proc. 49th Annu. Allerton Conf. Commun., Control, Comput.*, Monticello, IL, USA, Sep. 2011, pp. 1516–1520.

[22] R. Venkataramani, G. Kramer, and V. K. Goyal, “Multiple description coding with many channels,” *IEEE Trans. Inf. Theory*, vol. 49, no. 9, pp. 2106–2114, Sep. 2003.

[23] J. Pinkston, “An application of rate-distortion theory to a converse to the coding theorem,” *IEEE Trans. Inf. Theory*, vol. IT-15, no. 1, pp. 66–71, Jan. 1969.

[24] F. Jelinek, “Evaluation of distortion rate functions for low distortions,” *Proc. IEEE*, vol. 55, no. 11, pp. 2067–2068, Nov. 1967.

[25] A. D. Wyner, “The rate-distortion function for source coding with side information at the decoder–II: General sources,” *Inf. Control*, vol. 38, no. 1, pp. 60–80, Jul. 1978.

[26] W. H. R. Equitz and T. M. Cover, “Successive refinement of information,” *IEEE Trans. Inf. Theory*, vol. 37, no. 2, pp. 269–275, Mar. 1991.

[27] T. Berger, *Rate Distortion Theory: Mathematical Basis for Data Compression*. Englewood Cliffs, NJ, USA: Prentice-Hall, 1971.

[28] J.-J. Xiao and Z.-Q. Luo, “Compression of correlated Gaussian sources under individual distortion criteria,” in *Proc. 43rd Allerton Conf. Commun. Control Comput.*, Sep. 2005, pp. 438–447.

[29] J. Nayak, E. Tuncel, D. Gunduz, and E. Erkip, “Successive refinement of vector sources under individual distortion criteria,” *IEEE Trans. Inf. Theory*, vol. 56, no. 4, pp. 1769–1781, Apr. 2010.

[30] J. M. Bernardo, “The concept of exchangeability and its applications,” *Far East J. Math. Sci.*, vol. 4, pp. 111–122, 1996.

[31] W. Liu, G. Xu, and B. Chen, “The common information of N dependent random variables,” in *Proc. 48th Annu. Allerton Conf. Commun., Control, Comput.*, Monticello, IL, USA, Sep./Oct. 2010, pp. 836–843.

[32] G. Xu, W. Liu, and B. Chen, “Wyners common information for continuous random variables—A lossy source coding interpretation,” in *Proc. Annu. Conf. Inf. Sci. Syst.*, Baltimore, MD, USA, Mar. 2011, pp. 1–6.

[33] R. Tandon, L. Sankar, and H. V. Poor, “Multi-user privacy: The Gray–Wyner system and generalized common information,” in *Proc. IEEE Symp. Inf. Thoery*, St. Petersburg, Russia, Aug. 2011, pp. 563–567.

Ge Xu received the B.E. and M.E. degrees in Electrical Engineering from Xidian University, Xi’an, China in 2005 and 2008 respectively. She joined Syracuse University, Syracuse, NY in 2008 where she received the M.S. degree in Applied Statistics and the Ph.D. degree in Electrical and Computer Engineering in 2012 and 2013 respectively. Since 2014, she has been a research engineer in Nuance Communication. Her research interests include statistical signal processing and information theory.

Wei Liu received the B.E. degree in Electronic Engineering and Information Science from the University of Science and Technology of China (USTC), Hefei, China, in 2006, the M.S. degree in electrical engineering and the Ph.D. degree in electrical and computer engineering both from Syracuse University, Syracuse, NY, in 2009 and 2011, respectively. He is now with Bloomberg L.P., New York. His research interests include information security, joint source and channel coding, and network information theory.

Biao Chen (S’96–M’99–SM’07–F’15) received the B.E. and M.E. degrees in electrical engineering from Tsinghua University, Beijing, China, in 1992 and 1994, the M.S. degree in statistics and the Ph.D. degree in electrical engineering, from the University of Connecticut, Storrs, in 1998 and 1999, respectively. From 1994 to 1995, he worked at AT&T (China) Inc., Beijing, China. From 1999 to 2000, he was with Cornell University, Ithaca, NY, as a Postdoctoral Associate.

Since 2000, he has been with Syracuse University, Syracuse, NY, where he is currently a Professor in the Department of Electrical Engineering and Computer Science. His area of interest mainly focuses on signal processing and information theory for wireless communications and sensor networks.

Prof. Chen has served as Area Editor for the IEEE TRANSACTIONS ON SIGNAL PROCESSING, Associate Editor for the IEEE COMMUNICATIONS LETTERS, the IEEE TRANSACTIONS ON SIGNAL PROCESSING, and the *EURASIP Journal on Wireless Communications and Networking* (JWCN). He was a lead guest editor of a special issue on wireless sensor networks of the *EURASIP JWCN* and an elected member of the IEEE Signal Processing Society Sensor Array and Multi-Channel Signal Processing Technical Committee. He is currently a member of IEEE Signal Processing Society Signal Processing for Communication and Networking Technical Committee. He is the recipient of an NSF CAREER Award in 2006.